Riboswitch-mediated regulation of the *Mycobacterium tuberculosis* resuscitation promoting factor, *rpfB*

Stefan Schwenk

University College London

Submitted for the degree of Doctor of Philosophy

Declaration of authorship

I, Stefan Schwenk, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: Stefan Schwenk, _____

Abstract

Mycobacterium tuberculosis, the causative agent of tuberculosis, results in over 1.5 million deaths every year. An estimated third of the world's population may harbour a latent infection with risk of reactivation under permissive conditions. *M. tuberculosis* expresses a number of homologous resuscitation promoting factors (Rpfs) believed to facilitate resuscitation from dormancy. However, little is known about the regulation of Rpf expression. This study identifies a putative novel riboswitch in the 5'UTR of *rpfB*, which appears to be conserved in pathogenic mycobacteria, but not in non-pathogenic species. Northern blot analysis of *M. tuberculosis* RNA reveals short terminated transcripts and longer read-through transcripts suggesting transcription attenuation mediated by a transcriptional riboswitch. This has been further supported by 3'RACE, reporter gene assays and *in vitro* transcription. Findings indicate that *rpfB* is co-transcribed with the universally conserved rRNA methyltransferase *ksgA*, implicating the riboswitch in *ksgA* expression regulation, thereby linking resuscitation with ribosome maturation. Moreover, through translation start site mapping the RpfB start codon has been re-annotated and a ribosome binding site has been validated. This ribosome binding site is likely targeted by an *rpfB* antisense RNA produced from an antisense promoter that may also serve to regulate *rpfB* expression either by RNA polymerase clashing, sense-antisense interaction, mRNA processing or all of the above. Expression of the *rpfB* locus has been investigated under multiple physiological stresses and throughout biofilm formation. Collectively, findings support the presence of a novel *rpfB* riboswitch transcriptional 'on-switch' capable of regulating expression through transcription attenuation. Understanding the stimuli and regulation of Rpf gene expression will improve understanding of their role in disease and resuscitation.

Impact statement

Every year approximately 1.5 million people die as a result of *Mycobacterium tuberculosis* infection, making it the world leading cause of death due to a single infective agent (World Health Organisation, 2017). Whilst treatment is available, it requires a long 6-9 month multi-drug regimen. Many infections may also exhibit drug resistance that can result in treatment complications or failure. An estimated third of the world's population may harbour a latent infection; a non-active, asymptomatic, ostensibly non-infectious state. However, latently infected individuals may have an estimated 10% lifetime risk of progressing to active infection from latency. Strong evidence implicates this progression is attributed to the activity of resuscitation promoting factors (Rpfs). This study investigated a novel example of a type of regulatory RNA known as a riboswitch, which has the potential to regulate the resuscitation promoting factor, *rpfB*.

The work described in this thesis is the first instance of any such riboswitch having been discovered for *rpfB* regulation in any species. The novel riboswitch is conserved in pathogenic mycobacteria alone, where it regulates the expression of two genes, *rpfB* and *ksgA*, the later of which is involved in ribosome maturation for protein synthesis. Both of these genes contribute to processes well suited as targets for antibiotic therapies. As an impact of this study, in future the *rpfB* riboswitch could be targeted for the rational development of therapeutics with the potential to interfere with bacterial growth or dormancy, possibly in combination with conventional therapies. Theoretically this could help to reduce the reservoir of dormant *M. tuberculosis* that exists within latently infected individuals, with the goal of decreasing the number of deaths attributed to *M. tuberculosis* infection worldwide.

The research detailed within this thesis will have wider impacts for academia. The translational start site mapping method developed by Smollett *et al.* (2009) has been adapted, using a β-galactosidase reporter to re-annotate the *rpfB* protein coding sequence and identify a ribosome binding site. Such a method could be employed in future studies to investigate other potentially mis-annotated open reading frames in mycobacteria. These findings and the discovery of the *rpfB-ksgA* bi-cistron would very likely be reciprocated in other related actinobacteria that share sequence conservation.

The riboswitch has been further characterised through *in vitro* transcription assays using a system that has been developed and optimised throughout this study. The *M. tuberculosis* encoded *rpfB* riboswitch demonstrated unique interactions with recombinant *M. tuberculosis* RNA polymerase *in vitro*, likely due to the folding of nascent RNA. Besides highlighting the importance of using cognate systems to study riboswitches, the system that has been developed in this study will be used in future research to screen candidate ligands for the identification of the *rpfB* riboswitch cognate ligand.

Collectively the results of this study have been compiled into a manuscript for academic publication (Schwenk *et al.* 2018).

Acknowledgements

All *Mycobacterium tuberculosis* transformations and RNA extractions were carried out either by or with the assistance of Dr Alexandra Moores and Dr Kristine B. Arnvig.

Recombinant *Mycobacterium tuberculosis* RNAP holoenzyme was generously provided by Professor Sivaramesh Wigneshweraraj's group at the MRC Centre for Molecular Bacteriology and Infection, Imperial College London.

Bioinformatic analysis and sequence alignment of the *rpfB* locus was conducted by Dr Irilenia Nobeli, Birkbeck, University Of London to contribute to the publication of this work (Schwenk *et al.* 2018).

Dedication

Firstly, I would like to thank my primary supervisor, Dr Kristine B. Arnvig, for all of the advice, direction and support she has given me over the course of my research. Thank you for making me think, for making me better myself and for widening my expertise. You have taught me so much, which I shall not forget.

I would also like to thank Professor Timothy D. McHugh for being my secondary supervisor and so much more. You have helped me through difficult times and been a mentor to me my entire career so far. I cannot thank you enough for the kindness you have shown me.

Next I must thank my colleagues of the RNA(P) labs who have supported me and throughout my journey with their technical excellence and friendship. Special thanks to the denizens of RNApolynesia: Alex, Carol, Kathy and Gwenny. I have had the best time with you all, laughing, dancing, consuming emergency biscuits and injuring myself!

Thank you to my parents Gil and Elspeth, my brother Hans and sister Heidi, my parents in-law Mick and Linda, and to my closet friends who have shown considerable patience and understanding over these years. You have all given me encouragement and respite during my challenges.

Lastly, and most importantly, I thank my wife, Tara, to whom I dedicate this thesis. Thank you for believing in me. Through your support, advice and love you have earned this achievement just as much as I have. You have always been there, even before I started this journey, sharing in my successes and my failures. You have shouldered this burden with me in every way. I would not have succeeded without you.

Table of contents

Table of figures

Abbreviations

aa	amino acid
ADC	albumin dextrose catalase
ANOVA	analysis of variance
APS	ammonium persulfate
asRNA	antisense small RNA
bp	basepair(s)
BSA	bovine serum albumin
cDNA	complementary DNA
CDS	coding DNA sequence
CTD	C-terminal domain
DEPC	diethyl pyrocarbonate
dH_2O	deionised H_2O
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside
dsDNA	double stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
gDNA	genomic DNA
HIV	human immunodeficiency virus
K_d	dissociation constant
mRNA	messenger RNA
MU	Miller units
nt	nucleotide(s)
NTD	N-terminal domain
NTP	nucleoside triphosphate
OADC	oleic albumin dextrose catalase
OD	optical density
OPNG	ortho-nitrophenyl-β-galactoside
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PNK	polynucleotide kinase
RT-qPCR	real-time quantitative PCR
RACE	rapid amplification of cDNA ends
RBS	ribosomal binding site
RNA	ribonucleic acid
RNAP	RNA polymerase
rNTP	ribonucleoside
Rpf(s)	resuscitation promoting factor(s)
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	reverse transcriptase
RT-PCR	reverse transcription PCR

SAM	S-adenosyl-L-methionine
SD	Shine-Dalgarno sequence
SDM	site-directed mutagenesis
SDS	sodium dodecyl sulphate
SOC	super optimal broth with catabolite repression
sRNA	small RNA
SSC	salt sodium citrate buffer
ssDNA	single stranded DNA
TAE	Tris acetate-EDTA
TB	tuberculosis
TEC	transcription elongation complexes
TEMED	N,N,N',N-tetramethylethylenediamine
TEX	terminal exo-nuclease
TLSM	translational start site mapping
T_m	melting temperature
TNC	transiently non-culturable
Tris	tris(hydroxymethyl)aminomethane
TSS	transcription start site
TSSM	transcription start site mapping
UTR	un-translated region
UV	ultra violet
VBNC	viable but non-culturable
WT	wild type
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Nomenclature

The following nomenclature has been used to refer to DNA, RNA and protein:

- DNA – *rpfB*; italicised lowercase, uppercase italic for subtype
- RNA – rpfB; non-italicised lowercase, uppercase for subtype
- Protein – RpfB; non-italicised, uppercase first letter, uppercase subtype

All nucleic acid sequences are written in a 5' to 3' direction unless otherwise stated.

1 Introduction

1.1 Tuberculosis as a disease

Tuberculosis (TB) is the largest cause of death due to a single infective agent, *Mycobacterium tuberculosis* (World Health Organisation 2017). The *M. tuberculosis* acid-fast bacillus (see Figure 1.I) is an obligate pathogen that has successfully co-evolved alongside humans as a crowd disease with droplet-aerosol transmission (Comas *et al.* 2013). Treatment is available for tuberculosis as a 6-9 month multi-drug regimen (Bentrup & Russell 2001). The long time-course of treatment and frequency of side effects often result in non-compliance of treatment that can lead to exacerbated infection. This is compounded by increased prevalence of drug resistant and multi-drug resistant *M. tuberculosis*, as few alternative drugs are available. TB is more established in lesser-developed regions of the world where funding for treatment is limited, in addition there is often a stigma associated with the disease that prevents treatment being sought. In 2016 there were 6.3 million new cases of TB reported as part of a global estimated incidence of 10.4 million cases (see Figure 1.II).

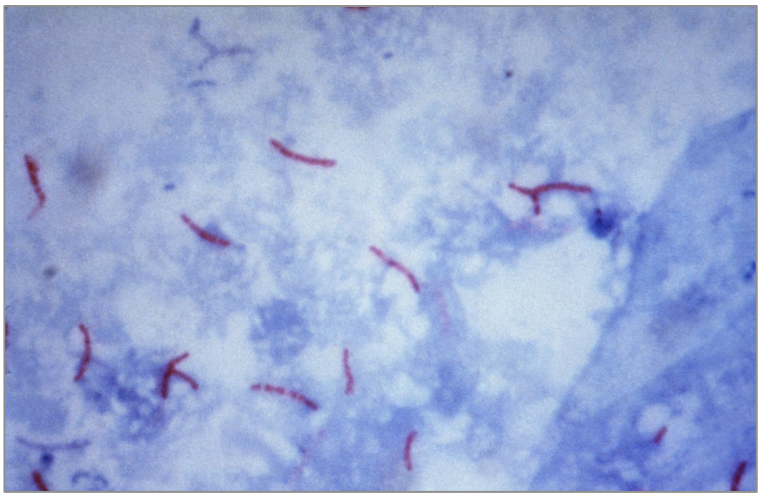


Figure 1.I photomicrograph of *M. tuberculosis*
Photomicrograph shows acid-fast *M. tuberculosis* bacilli visualised using Ziehl-Neelsen staining; carbol fuchsin stain is retained in the mycolic acids of the mycobacterial cell wall during acid-alcohol de-staining (acid-fast). Image taken under 1000x magnification (Centers for Disease Control and Prevention & Kubica 1979).

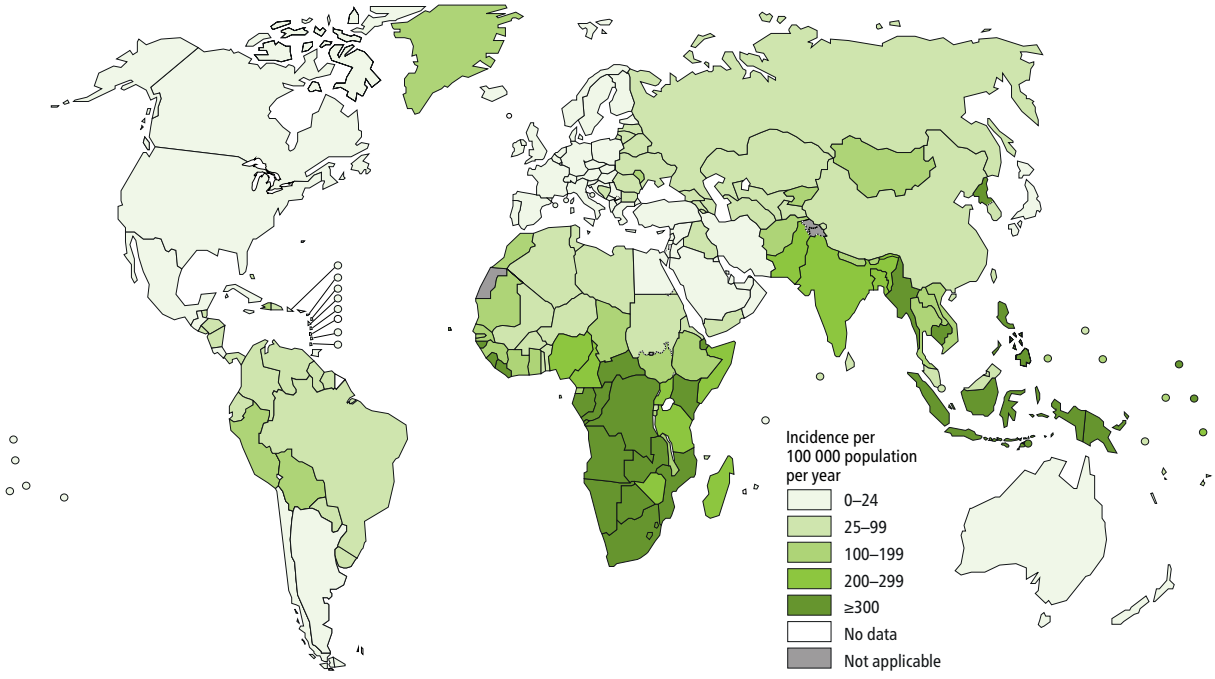


Figure 1.II estimated global TB incidence in 2016
Diagram illustrates the global TB incidence in 2016, estimated at 10.4 million new cases by the World Health Organisation Global (World Health Organisation 2017).

An estimated third of the world's population is either actively or latently infected with the pathogen (Bentrup & Russell 2001). Latent infection refers to a non-active, asymptomatic, ostensibly non-infectious TB infection with a 10% lifetime risk of progressing activation (10% per year for those co-infected with HIV) (Corbett *et al.* 2003). Latently infected individuals act as a reservoir of the disease, ensuring continuity of the pathogen and dissemination within the host population once activated (Comas *et al.* 2013).

1.2 Dormancy in mycobacteria

In natural environments bacteria can adopt 'non-growth' states to resist stress. For example, viable but non-culturable (VBNC) bacteria exhibit very low metabolic and reproductive activity, although live/dead staining indicates they are still alive. The vast majority of VBNC bacteria cannot be grown under standard laboratory conditions, however some may be resuscitated to culturability by specific growth temperatures or with supplementation with culture filtrate from growing cells (Oliver 2005).

M. tuberculosis may persist within granulomas >30 years before reactivating to cause disease once again (Lillebaek *et al.* 2002). It is therefore more appropriate to refer to *M. tuberculosis* as transiently non-culturable (TNC) because it does not necessarily remain non-culturable (Mukamolova *et al.* 2003). It is not currently understood exactly what conditions are permissive for mycobacterial resuscitation, but strong evidence implicates the activity of resuscitation promoting factors (Rpfs).

1.3 Resuscitation promoting factors

Rpfs were originally identified in *Micrococcus luteus* (Mukamolova *et al.* 1998). *M. luteus* enters a dormant state after growth to stationary phase and starvation in nutritionally depleted media. However, resuscitation occurs upon supplementation with filtered culture supernatant from actively growing cells. Subsequent studies have revealed this is due to the activity of a cell wall lytic enzyme named 'Rpf' for which homologues exist in other G:C rich bacteria, particularly mycobacteria (Mukamolova, Turapov, Young, *et al.* 2002; Mukamolova *et al.* 2006).

M. tuberculosis encodes five resuscitation promoting factors, *rpf*(*A*-*E*), which are functionally redundant of each other to a certain extent (Zhonghe & Zhang 1999; Tufariello,

Jacobs & Chan 2004; Downing *et al.* 2005; Tufariello *et al.* 2006). Currently there is little known about the regulation of *rpf* expression in *M. tuberculosis*, although findings do indicate that regulation occurs in response to various stresses (Rickman *et al.* 2005; Gupta, Srivastava & Srivastava 2010). There is evidence that *M. tuberculosis* Rpfs may also be linked to virulence as mutants lacking multiple *rpf* genes lose virulence (Kana *et al.* 2008; Kondratieva *et al.* 2011). Understanding the mechanisms and conditions through which *rpf* gene expression is achieved may improve understanding of their role in disease and resuscitation settings.

1.4 Transcription regulation

The central dogma of molecular biology describes a flow of genetic information from DNA to RNA to protein. Transcription is the process by which DNA is copied to RNA by the multi-subunit enzyme RNA polymerase (RNAP). The cycle of transcription initiation, elongation and termination is one of the earliest points at which regulation of gene expression may be modulated.

1.4.1 Transcription initiation

The prokaryotic RNAP consists of two large subunits (β and β') along with two alpha (α) subunits which when combined are stabilised by an omega (ω) subunit to form the RNAP core enzyme ($\beta\beta'\alpha_2\omega$). Core RNAP may interact with a number of sigma (σ) factors to form a holoenzyme capable of binding DNA at promoter regions (Gusarov & Nudler 1999; Ruff, Record & Artsimovitch 2015). In prokaryotes, promoter specificity is achieved almost entirely through sigma factor recognition (Sachdeva *et al.* 2010). The hexameric -10 and to a lesser extent -35 promoter elements contain consensus sequences which are recognised by specific sigma factors. Transcription of essential 'housekeeping' genes is regulated by at least one sigma factor with additional sigma factors regulating genes in response to specific environmental changes or physiological stresses. Species that are subject to more variable environments or those that have larger genomes tend to have a greater variety of sigma factors than others.

M. tuberculosis is subject to multiple stresses and environments through the course of transmission, infection, immune response attack, dormancy and resuscitation. To adapt to

these changing environments *M. tuberculosis* has thirteen known sigma factors, a larger number in relation to its genome size when compared with other pathogens (Sachdeva *et al.* 2010). SigA (σ^A) executes basal transcription of 'housekeeping' genes whilst other factors respond to stresses and environmental changes. Transcription initiation can also be further regulated through binding of promoter proximal repressors or activators.

1.4.2 Transcription termination

Subsequent to transcription initiation, the polymerase dissociates from the sigma factor and escapes the promoter region as a transcription elongation complex (TEC). The TEC progresses along the DNA template, maintaining a transcription bubble of 12-14 nucleotides (nt) of single stranded DNA (ssDNA) inside of which a ~7-9 nt DNA:RNA hybrid exists between the template DNA and nascent RNA being synthesised (Darst 2001; Peters, Vangeloff & Landick 2011). Maintenance of the transcription bubble, DNA:RNA hybrid and extrusion of the nascent RNA are critical to the stability and processivity of the TEC, affecting the rate of RNA polymerisation as well as the frequency with which RNAP dissociates from DNA. Transcription obstacles like repeating nucleotide motifs or DNA damage may slow down, pause or arrest the TEC. The stability and processivity of the complex may also be manipulated through the binding of additional transcription factors, such as NusA and NusG which enhance pausing and termination (Artsimovitch & Landick 2000; Czyz *et al.* 2014). Arrested complexes may be rescued by backtracking, where nascent RNA is extruded from RNAP through a secondary exit channel and cleaved so that transcription may be re-initiated or damage repaired (Tetone *et al.* 2017). Premature transcription termination results in the loss of incomplete transcripts. This irreversible event is deleterious to successful RNA transcription as incomplete transcripts cannot be re-initiated. TECs that cannot be re-initiated will be terminated by either a factor dependent or independent mechanism.

Factor-dependent termination relies on the binding of Rho protein to C-rich '*rut* sites' as a homohexameric ring. The nascent RNA is then translocated in a 5' to 3' direction through the ring, tautening the RNA to destabilise the DNA:RNA hybrid of the TEC (Richardson 2002). Rho may also enhance transcription termination when ribosomes are coupled to the TEC by binding with NusG C-terminal domain (CTD) (Peters, Vangeloff & Landick 2011).

Factor-independent termination or intrinsic termination is facilitated by the formation of a G:C rich RNA stem-loop immediately followed by a 7-9 nt poly-uracil tract (Czyz *et al.* 2014). The stem-loop poly(U) structure destabilises the TEC to cause transcription termination. After the TEC has transcribed the G:C rich stem-loop sequence it reaches the poly(U) tract. Here the weak A:U basepair interactions that form between DNA:RNA hybrid at the active site cause pausing of the elongation complex (Gusarov & Nudler 1999). This pausing allows time for stem-loop nucleation which, depending on the sequence identity, may fold within microseconds of exiting the polymerase (Peters, Vangeloff & Landick 2011). The stem folds until it physically pushes against RNAP causing both a deformation of the RNAP exit channel as well as translocation of the nascent RNA from the active site, melting the DNA:RNA hybrid within the complex (Gusarov & Nudler 1999; Czyz *et al.* 2014). It is possible that the TEC may attempt backtracking once paused at the poly(U) site, however stem-loop formation would make it impossible for the TEC to backtrack >1-2 nt, effectively trapping the TEC (Gusarov & Nudler 1999). Unable to backtrack and lacking contacts with the DNA:RNA hybrid, the paused TEC tends much more favourably towards dissociation thus terminating transcription.

Mycobacterial TECs may terminate at imperfect poly(U) tracts of ≥4U residues, even if subsequent nucleotides contain other bases, an effect known to be enhanced by binding of NusG N-terminal domain (NTD) (Czyz *et al.* 2014). NusG-NTD binds RNAP, stabilising its clamp on the DNA to increase processivity, therefore making it easier for hypertranslocation of the RNA from the DNA:RNA hybrid during stem-loop formation. Conversely, NusA can decrease processivity and increase intrinsic termination by prolonging the time with which paused TECs dwell at the poly(U) tract, allowing a greater timeframe for stem-loop nucleation (Ha *et al.* 2010).

1.5 Translation regulation

Translation is the synthesis of polypeptide using transcribed messenger RNA (mRNA) as a template. The process of translation may also be manipulated to modulate gene expression. Prokaryotic translation is coupled with transcription such that translation initiation occurs on mRNA transcripts as they are being transcribed. Briefly, during canonical translation initiation the ribosome associates with the mRNA, binding at a region complementary to the

16S ribosomal RNA (rRNA) called the Shine-Dalgarno (SD) sequence (Shine & Dalgarno 1975; Laursen *et al.* 2005). The SD sequence constitutes a ribosomal binding site (RBS) where the small 30S and large 50S ribosomal subunits bind to assemble the full 70S ribosome. Coupled to the RNAP by NusG-CTD, the ribosome commences protein synthesis from the translation start codon that marks the start of the coding region (Peters, Vangeloff & Landick 2011). The mRNA transcript upstream from the start codon is known as the 5'un-translated region (UTR). Translation elongation continues from the start codon to synthesise the polypeptide, terminating translation at the stop codon. The region between the start codon and stop codon of DNA that encodes the polypeptide sequence is known as the open reading frame (ORF).

Mycobacteria are capable of initiating translation from ATG (AUG in mRNA) as well as uncommon start codons such as GTG (GUG) just as often and more rarely TTG (UUG); although other prokaryotes may also utilise uncommon start codons, mycobacteria do so far more frequently (DeJesus, Sacchettini & Ioerger 2013). This is often a cause of incorrect annotation of ORFs in mycobacterial genomes, which are often annotated by identification of RBS or start codon (Smollett *et al.* 2009; DeJesus, Sacchettini & Ioerger 2013). In addition, many mycobacterial genes are leaderless and do not contain a 5'UTR, therefore lacking a canonical RBS (Cortes *et al.* 2013).

1.6 Regulatory RNA

Contrary to the central dogma of molecular biology, not all RNA encodes protein. Some non-coding RNAs may act as regulators of transcription and translation or influence the stability of other RNA transcripts. Two major types of regulatory RNAs are those of small RNAs (sRNAs) and riboswitches. There are also clustered regularly interspaced short palindromic repeat (CRISPR) RNAs used by prokaryotes during phage infection, but these are beyond the scope of the present study. Regulatory RNAs are important molecules utilised to modulate gene expression at the transcriptional and post-transcriptional level.

Small RNAs are relatively short non-coding transcripts, generally ~55-150 nt in length (Wagner, Altuvia & Romby 2002). Though some protein binding sRNAs are known, most basepair in a complementary fashion with other nucleic acids (Arnvig & Young 2009). They

are transcribed much like mRNA transcripts either *cis* or *trans* to their targets, yielding molecules with diverse regulatory functions often associated with stress adaptation.

1.6.1 *Trans*-encoded sRNA

Trans-encoded sRNAs are not transcribed from their target-binding region but rather from chromosomally distant locations (Waters & Storz 2009). As such, *trans*-sRNAs may bind multiple targets using limited sequence complementarity of ~10-25 nt with a core seed region. *Trans*-sRNA binding typically down-regulates gene expression by occluding the RBS and/or causing RNA instability through degradation (Papenfort & Vogel 2010). Alternatively, binding may serve to relieve secondary structures that occlude the SD sequence, permitting translation initiation. Many *trans*-sRNAs serve to regulate genes involved in stress response and virulence, responding to stimuli to co-ordinate regulation of multiple genes either through direct basepairing or modulation of transcription factors and activators (Waters & Storz 2009; Papenfort & Vogel 2010).

Due to their limited sequence complementarity, *trans*-sRNA binding is often facilitated by the chaperone protein Hfq. Characterised in *Escherichia coli*, this homohexameric ring has orthologues in both Gram positive and negative bacteria (Storz, Opdyke & Zhang 2004; Brennan & Link 2007). Hfq provides a platform onto which mRNAs and sRNAs may assemble, melting secondary structures in the RNA. Binding also appears to increase sRNA stability, as Hfq binding sites are similar to those of RNase E, protecting the sRNAs from enzymatic degradation (Storz, Opdyke & Zhang 2004). Conversely, it has been suggested that once an sRNA binds its target Hfq may also recruit the degradosome to degrade both the sRNA and bound target (Gottesman 2005; Waters & Storz 2009). Deletion of Hfq attenuates virulence in many bacterial pathogens (Papenfort & Vogel 2010).

Due to its abundance and known association, the use of co-immunoprecipitation of Hfq has aided identification of numerous sRNAs (Zhang *et al.* 2003). However, some bacterial phyla such as actinobacteria do not encode any known Hfq homologue (Storz, Opdyke & Zhang 2004). *M. tuberculosis* has no known homologue; although candidates have been suggested, none are expressed at similar concentrations to Hfq in *E. coli* (Arnvig & Young 2012). Due to the high G:C content (67%) of *M. tuberculosis* DNA it may be that *M. tuberculosis* *trans*-sRNAs are capable of binding without chaperone assistance as the intermolecular and

intramolecular forces of G:C basepairing allow stronger annealing of RNA duplexes than A:U basepairing.

1.6.2 *Cis*-encoded sRNA

Cis-encoded sRNAs are transcribed from the antisense strand opposite their targets, often being referred to as antisense small RNAs (asRNAs) (Brantl 2007). asRNAs regulate gene expression through a variety of mechanisms: binding may induce folding of target RNA so that an intrinsic terminator forms to facilitate premature termination, alternatively binding may occlude the SD sequence of mRNA or interfere with folding required to reveal the SD sequence to inhibit translation initiation (Gottesman 2005; Brantl 2007). Sequence complementarity between asRNA and target enables RNA duplex formation, which is recognised and cleaved as RNase III substrate if >20 basepairs (bp) double stranded RNA (Brantl 2007; Arnvig & Young 2009). Conversely, binding may stabilise mRNA transcripts to prevent their degradation (Brantl 2007). The majority of asRNAs are associated with prokaryotic mobile genetic elements such as plasmids, phages and transposons; however those that are chromosomally encoded tend not to be *cis*-acting but instead act *trans* to an array of targets (Wagner, Altuvia & Romby 2002). Nevertheless, there are exceptions with chromosomally encoded asRNAs being discovered more recently through total RNA transcriptomics (Brantl 2012).

Cis-encoded elements or *cis*-acting elements that form within the coding mRNA 5'UTR may sense stimuli, changing folding conformation to sequester or reveal the SD sequence to modulate translation or form intrinsic terminators to modulate transcription (Papenfort & Vogel 2010). One notable example is that of 'RNA thermometers'. These regulatory elements use specific physiological temperatures to melt secondary structures within the 5'UTR to reveal the SD sequence for ribosome binding, thus exerting translational control (Johansson *et al.* 2002). Such conformational changes in response to stimuli are also observed in a more complex type of *cis*-encoded elements known as riboswitches.

1.7 Riboswitches

First coined in 2002, the term riboswitch refers to regulatory elements of RNA that are mostly situated within the 5'UTR of mRNA transcripts (Nahvi *et al.* 2002). Capable of binding cognate ligands with high specificity and affinity, riboswitches modulate expression of downstream coding regions through conformational changes (Nahvi *et al.* 2002; Roth & Breaker 2009). Examples of riboswitches have been found throughout eubacteria along with limited incidence in eukaryotes, suggesting an evolutionary conserved origin (Loh *et al.* 2009; Roth & Breaker 2009).

1.7.1 Riboswitch architecture

Riboswitches are generally thought to be composed of two domains: an aptamer domain that binds target ligands and an expression platform that is manipulated by ligand binding events to affect regulation of gene expression via RNA conformational changes (Tucker & Breaker 2005). Expression platforms are highly specific for their associated ligands, many of which are products of the regulated genes themselves thereby acting as feedback regulators (Nudler & Mironov 2004).

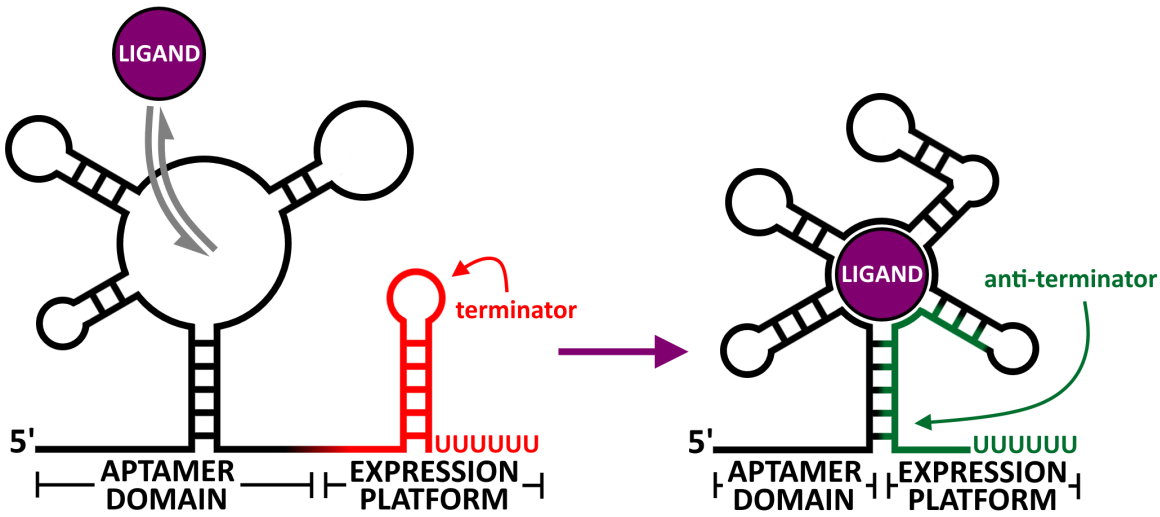


Figure 1.III example of riboswitch architecture
An example of a transcriptional riboswitch: in this example when ligand is not bound to the aptamer domain the expression platform adopts the structure of an intrinsic terminator resulting in transcription termination (left). Ligand binding in equilibrium results in a structural remodelling to form an anti-terminator permissive of transcriptional read-through (right) as a transcriptional 'on-switch'.

The majority of riboswitches employ one of two expression platforms, although there are less common examples of ribozyme self-cleavage expression platforms that render transcripts vulnerable to RNase degradation (Winkler *et al.* 2004). The first main type of expression platform is that of the intrinsic terminator, a stem-loop abutted by a poly(U) tract (see Figure 1.III). The default conformation of the riboswitch in absence of bound ligand may be either a terminator or anti-terminator. Upon ligand binding to the aptamer domain the alternative conformation is induced in the expression platform, a terminator conformation would become an anti-terminator and *vice versa*. In this regard the intrinsic terminator expression platform modulates transcriptional read-through into the coding region of mRNA in response to ligand binding.

The second type of expression platform is that of SD sequence sequestration. Here an anti-Shine-Dalgarno sequence basepairs with the downstream SD sequence in the 5'UTR, or the secondary structure of the RNA is such that the SD sequence becomes sequestered, inhibiting translation. The default state may be that of a sequestered or accessible SD sequence. Upon ligand binding, the alternative conformation is adopted, such that the sequestered SD sequence becomes accessible to ribosome binding and *vice versa*. This mechanism allows the riboswitch to modulate translational attenuation in response to ligand binding.

The architecture of riboswitches is generally a single specific aptamer domain coupled with a single expression platform, though there are variations. Tandems of aptamer domains or complete riboswitches specific for single or multiple ligands can be combined with tandem expression platforms (Roth & Breaker 2009). This allows highly specialised genetic regulation capable of integrating multiple stimuli for multiple expression outputs. Recently a flavin mononucleotide (FMN) sensing riboswitch was found in *Escherichia coli*, capable of utilising a single aptamer domain with both an intrinsic terminator and SD sequence sequestration expression platform, achieving maximum gene down-regulation from a single riboswitch (Pedrolli *et al.* 2015). This modularity of riboswitches makes them powerful regulators of gene expression at both the transcriptional and post-translational level, capable of executing regulation in response to very specific stimuli.

1.7.2 Riboswitch switching mechanisms

Any given riboswitch will have a thermodynamically favourable default conformation that is altered either in the presence or absence of ligand. The ligand concentration threshold at which this switching event occurs is dictated by a thermodynamic equilibrium influenced by a dissociation constant (K_d), "the ratio of the concentrations of free ligand and aptamer to that of the aptamer-ligand complex, measured at equilibrium" (Roth & Breaker 2009). Therefore if a sufficient concentration of ligand is available for binding – as dictated by the K_d – the thermodynamic equilibrium will shift in favour of the alternative conformation. This structural change will then be commuted through to the associated expression platform to affect a change in gene expression. This makes riboswitch regulated gene expression highly sensitive to the presence or absence of associated ligands; some may switch only within specific concentration physiologically thresholds (Lutz *et al.* 2014).

Notwithstanding, ligand binding dictated by thermodynamic equilibrium may not always be the most important factor for switching events. As transcription occurs in a 5' to 3' direction the aptamer domain of a riboswitch is synthesised first, upstream of the expression platform. There is a limited timeframe in which the aptamer may adopt a secondary structure, bind its cognate ligand and commute the binding event through to the expression platform. If the expression platform folds before the thermodynamic equilibrium shifts then the genetic decision may be inevitable, especially if the equilibrium is slow to equate (Wickiser *et al.* 2005). This is particularly true for intrinsic terminator expression platforms, which require the poly(U) tract to be within the RNAP active site. In this regard some riboswitches may sometimes operate under kinetic parameters determined by transcription elongation rate or ligand binding kinetics (Wickiser *et al.* 2005).

1.7.3 *Trans*-acting riboswitches

It is not unprecedented that riboswitches may regulate other genes in *trans*. Riboswitches that utilise intrinsic terminators may yield numerous short attenuated transcripts that are very similar to sRNAs. Research has revealed that these short transcripts may have the capacity to function as *trans*-sRNAs, basepairing with other transcripts to modulate expression of genes in addition to their *cis*-acting regulatory capacity (Loh *et al.* 2009). This demonstrates that riboswitches may have a wider impact on gene expression. Several

riboswitches have been discovered that do not control expression of coding genes but rather transcription of regulatory sRNAs. These transcripts then affect the expression of multiple genes, acting in *trans* as sRNAs or by sequestering regulatory proteins (DebRoy *et al.* 2014; Mellin *et al.* 2014; Mellin & Cossart 2015).

1.8 The *rpfB* locus

RpfB (Rv1009) is a cell wall hydrolase, one of the five Rpfs expressed by *M. tuberculosis* (Mukamolova *et al.* 2006). Evidence suggests that RpfB may be both membrane lipoprotein anchored as well as secreted to act on neighbouring cell walls (Downing *et al.* 2004). RpfB has been found to interact with a second cell wall hydrolase, RipA, for increased cell wall digestion and resuscitation activity (Hett *et al.* 2007, 2008). It is has been postulated that activity on cell wall peptidoglycan may be key in its resuscitating activity, with muropeptide digests acting as possible signalling molecules (Keep *et al.* 2006). Global transcription start site mapping (TSSM) of *M. tuberculosis* indicates that there are large 5'UTR upstream regions to some *rpf* mRNAs (Cortes *et al.* 2013). Two transcription start sites (TSSs) have been identified upstream of *rpfB* (see Figure 1.IV) and used to annotate two putative promoters designated P_1 and P_2 (Cortes *et al.* 2013). In addition there appears to be signal corresponding to an antisense transcript and promoter designated P_{as}.

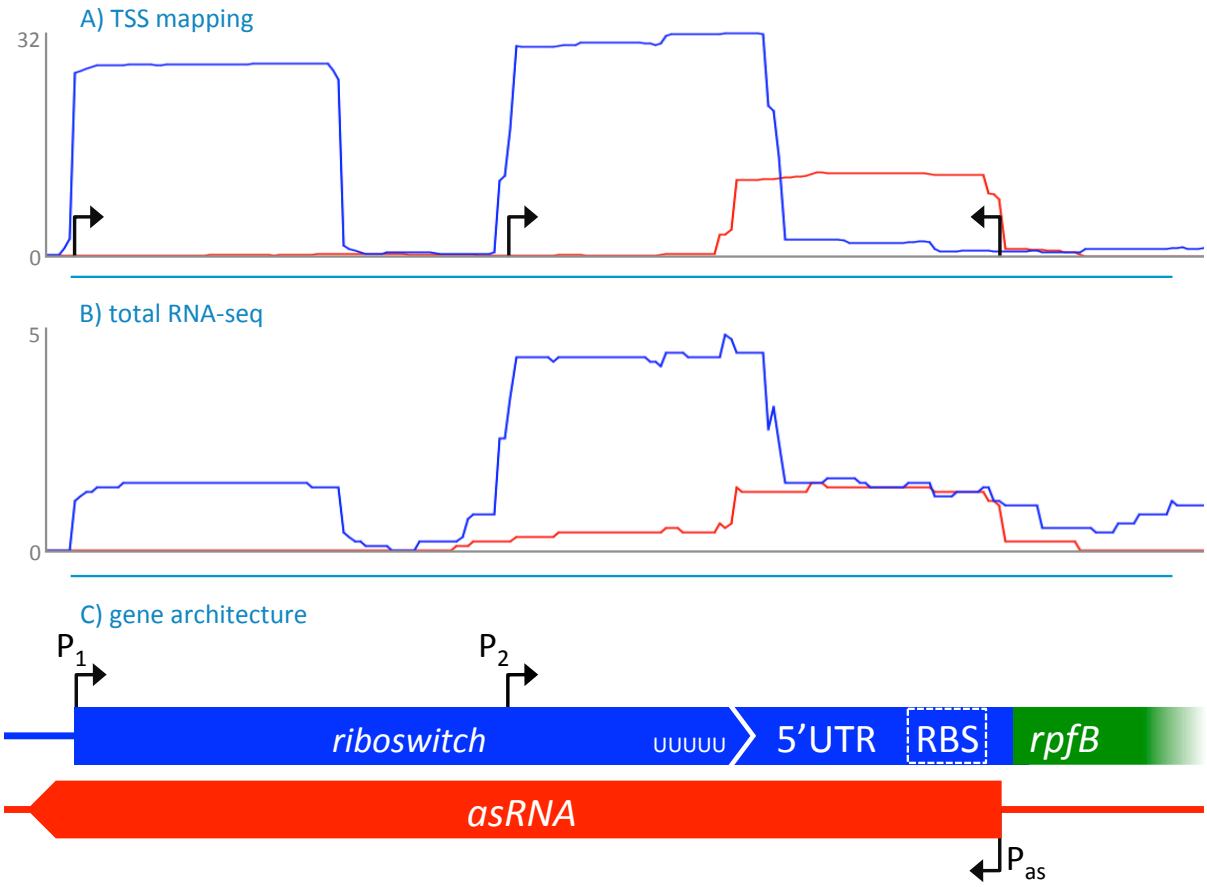


Figure 1.IV diagram of *rpfB* RNA expression
Chart (A) shows global transcriptional start site analysis of sense (blue) and antisense (red) primary transcript abundance in *M. tuberculosis* (Cortes *et al.* 2013). Chart (B) shows the RNA-seq transcript abundance of sense (blue) and antisense (red) transcripts normalised to total reads (Cortes *et al.* 2013). Diagram (C) shows the organisation of genetic elements in alignment with the charts above it. Black arrows indicate putative transcription start sites.

1.9 The *rpfB* riboswitch

Transcripts initiated at the *rpfB* P_1 TSS may be expected to yield a large 5'UTR characteristic of a *cis*-acting regulatory RNA, encoding a potential intrinsic terminator. Since a terminator would prevent further transcription into the *rpfB* coding sequence (CDS), it is possible it may constitute the expression platform of a transcriptional riboswitch. Riboswitch regulation of Rpfs is not unprecedented as a riboswitch has already been identified in the 5'UTR of *rpfA* (Block, Hammond & Breaker 2010; Arnvig & Young 2012; Nelson *et al.* 2013; St-Onge *et al.* 2015). Structural modelling (Zuker 2003) of the *rpfB* riboswitch candidate region (which shall forthwith be referred to as a riboswitch) predicts only a single structure that resembles an intrinsic terminator (see Figure 1.V). Whilst the predicted structure does resemble a

stem-loop poly(U) tract, the sequence of consecutive uracil residues would likely be insufficient for recognition as an intrinsic terminator (Czyz *et al.* 2014). Moreover, a lack of any predicted alternative structure would make it unlikely the riboswitch region could adopt conformational changes in response to ligand binding, a prerequisite of riboswitch function.

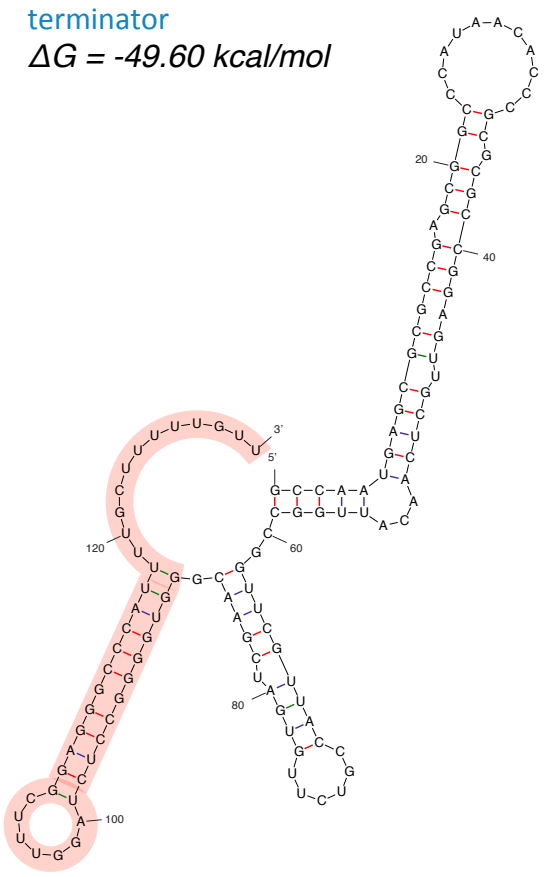


Figure 1.V unconstrained *rpfB* riboswitch predicted structure
Structural prediction of the *rpfB* riboswitch generated using Mfold (Zuker 2003). The proposed intrinsic terminator structure is highlighted in red.

During intrinsic termination, the poly(U) tract of an intrinsic terminator is precluded from secondary structure formation as it is bound as part of the DNA:RNA hybrid within the RNAP active site. Applying a single stranded basepairing constraint to the *rpfB* riboswitch poly(U) tract for structural modelling results in only two predicted structures: a terminator conformation and an anti-terminator conformation (see Figure 1.VI). The predicted terminator conformation displays a poly(U) tract more representative of canonical intrinsic terminator with four consecutive uracil immediately downstream of the stem-loop. Comparatively, the predicted free energy of these structures demonstrates the terminating conformation is more thermodynamically stable compared to the anti-terminating

conformation, suggesting that the terminating state is the default state. Therefore the *rpfB* riboswitch likely represents a rarer example of a transcriptional 'on-switch' where the less favourable anti-terminator or read-through conformation may be adopted through ligand interactions (Tucker & Breaker 2005).

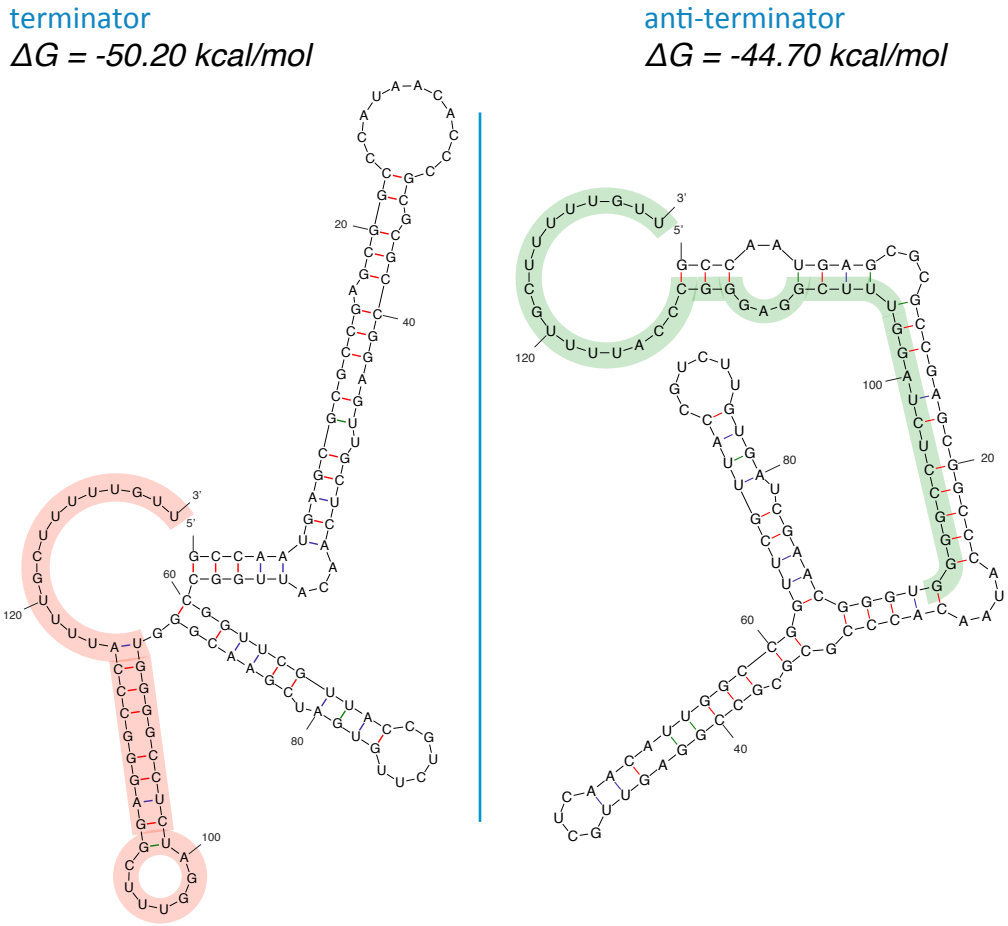


Figure 1.VI *rpfB* riboswitch predicted structures with constrained poly(U) tract
Structural predictions of the *rpfB* riboswitch generated using Mfold (Zuker 2003) with a single stranded basepairing constraint applied to the poly(U) tract. Terminator (left) and anti-terminator (right) conformations with identical red and green sequences highlighted for comparison.

Multiple sequence alignment of the *rpfB* upstream region across mycobacteria reveals the P_2 -10 promoter element is well conserved in numerous species (see Figure 1.VII). Conversely the P_1 -10 promoter element and downstream riboswitch sequence appears to be conserved in pathogenic mycobacterial species alone, principally the *M. tuberculosis* complex (Schwenk *et al.* 2018). Therefore it seems the proposed *rpfB* riboswitch may represent a novel riboswitch in mycobacteria, potentially associated with pathogenesis.

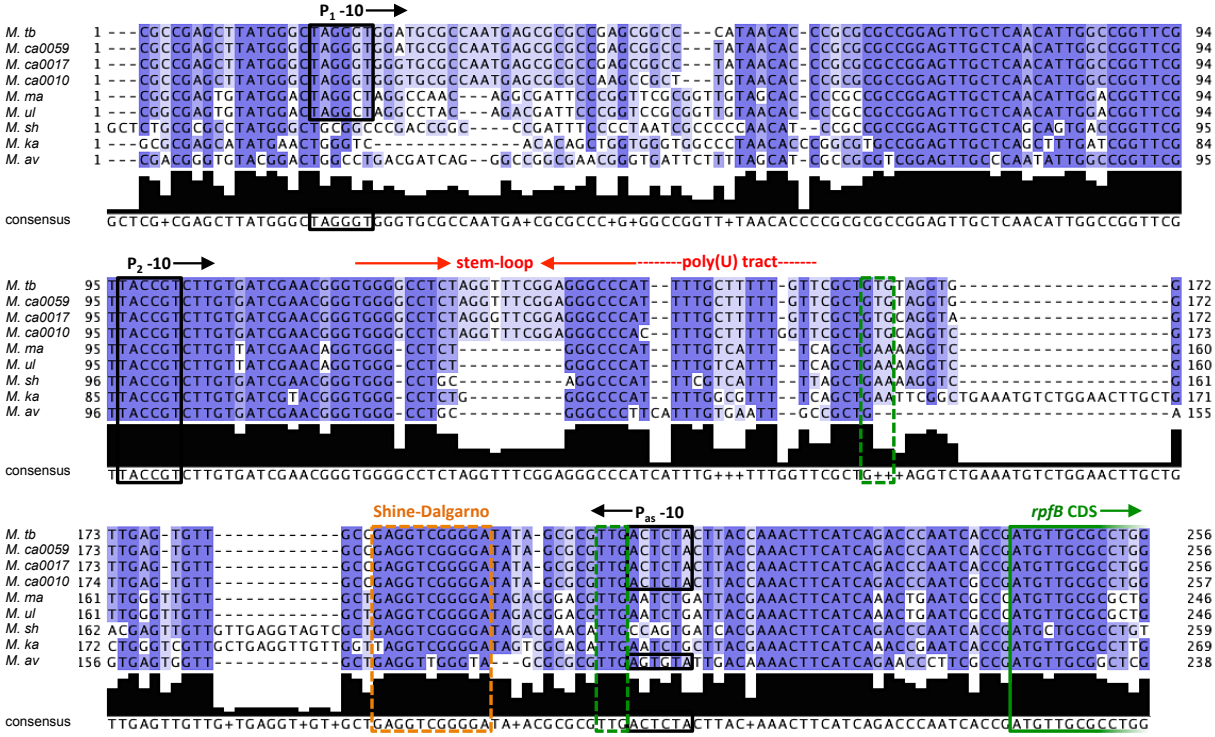


Figure 1.VII sequence alignment of *rpfB* promoter regions and 5'UTR
Sequence alignment of *M. tuberculosis* (*M. tb*), three strains of *M. canettii*: CIPT 140010059 (*M. ca0059*), CIPT 140070017 (*M. ca0017*) and CIPT 140070010 (*M. ca0010*), *Mycobacterium marinum* (*M. ma*), *Mycobacterium ulcerans* (*M. ul*), *Mycobacterium shinjukuense* (*M. sh*), *Mycobacterium kansasii* (*M. ka*) and *Mycobacterium avium* (*M. av*) constructed in Jalview. Adapted from (Schwenk *et al.* 2018). Residues coloured by percentage sequence conservation (darker blue indicates higher conservation). Putative P_1, P_2 and P_{as} -10 promoter elements are highlighted in black. First four codons of *rpfB* CDS highlighted in green, with alternative start codons highlighted with dashed lines. Proposed SD sequence is highlighted in orange (see section 3.4). Red annotations denote the stem-loop poly(U) tract of the proposed riboswitch intrinsic terminator identified in *M. tuberculosis*.

1.10 Project objectives

The present study was based on the observation that *M. tuberculosis rpfB* locus encodes a putative riboswitch region, potential regulatory asRNA and that these regulatory RNA elements may modulate the expression of *rpfB* at the transcriptional and post-transcriptional level.

1.10.1 Mapping of the *rpfB* translational start site

Correct annotation of translational start sites can be achieved through translational start site mapping (TLSM) (Smollett *et al.* 2009). *rpfB* has three potential translational start sites, most upstream of which is GTG (GUG in mRNA) followed by TTG (UUG) and finally the annotated ATG (AUG) start codon (see Figure 1.VII). This annotation had not been experimentally verified until recently when Sharma *et al.* (2015) reported a significant

decrease in *rpfB-lacZ* translational reporter β-galactosidase expression on mutation of the ATG start codon to AAG. However, no other possible start sites were investigated.

In the *Mycobacterium smegmatis rpfB* homologue (MSMEG_5439) the annotated start codon is further upstream, proximal to the TTG rather than ATG compared to *M. tuberculosis* (see Appendix Figure 10.I). In *M. tuberculosis* there is no apparent ribosome binding site within a feasible distance upstream of the *rpfB* ATG start codon. If the CDS were extended upstream to the TTG start codon it would be more closely situated to a potential RBS (see Figure 1.VII). Moreover, the CDS 5'terminus would be closer to the 5'end of the proposed asRNA, a location more complementary of canonical asRNA function, potentially serving to occlude the RBS. Therefore the project aimed to determine the translational start site of *rpfB* through TLSM, to verify the Sharma *et al*. (2015) ATG annotation (see section 3) and investigate the additional possible upstream start codons. This research also served to enable correct construction of transcriptional and translational reporter constructs.

1.10.2 A potential link between *rpfB* and *ksgA*

In *M. tuberculosis* there is a 28 bp overlap between the 3'end of the *rpfB* CDS and the 5'end of a gene named *ksgA*, which encodes a universally conserved yet non-essential methyltransferase required for functional ribosome biogenesis (Mangat & Brown 2008). This proximity of cell wall enzymes with a *ksgA* methyltransferase is extensively conserved in many firmicutes and actinobacteria (Ravagnani, Finan & Young 2005). Inspection of transcription start site mapping upstream of *ksgA* indicates a lack of a TSS for the gene (Cortes *et al.* 2013). It is therefore possible that *ksgA* is co-transcribed from the upstream *rpfB* P_1 or P_2 promoters, potentially linking resuscitation from dormancy by *rpfB* with ribosome maturation by *ksgA***.** Consequently this project aimed to explore any potential regulatory or functional link between *rpfB* and *ksgA* at the transcriptional and translational level.

1.10.3 Characterisation of the *rpfB* riboswitch

This project aimed to characterise the putative *rpfB* riboswitch with the potential to elucidate a previously un-described post-transcriptional regulation of *rpfB* in *M. tuberculosis*. Research objectives included: ascertaining potential mechanisms of riboswitch-mediated attenuation, exploring conditions permissive of riboswitch expression and regulation, and assaying riboswitch-mediated attenuation both *in vitro* and *in vivo* to aid the identification and validation of candidate ligands.

1.10.4 Understanding the *rpfB* transcriptional unit

In addition to the putative P_1 and P_2 promoters identified through transcription start site mapping (Cortes *et al.* 2013), recently an additional promoter was described upstream of an alternative *rpfB* TSS. Sharma *et al.* (2015) described a SigB consensus -10 promoter element downstream of the P_1 and P_2 promoters. Transcripts promoted by the proposed SigB element would not encode the potential *rpfB* riboswitch and therefore preclude the possibility of riboswitch-mediated regulation of downstream genes. Therefore this project aimed to understand the *rpfB* transcriptional unit by characterising putative promoter elements and clarifying conflicting reports.

In addition to the identified sense stand promoters; the RNA-seq data by Cortes *et al.* (2013) indicates the possibility of an asRNA and associated promoter. This asRNA may allow another layer of regulation to the *rpfB* locus either through occlusion of the RBS, transcription down-regulation through polymerase clashing or transcript degradation as regulatory asRNA. Therefore the project also aimed to characterise the proposed P_{as} promoter and possible asRNA transcript regulatory function.

2 Materials and methods

2.1 Culture media

2.1.1 Luria-Bertani (LB) broth

1 L LB broth was made by dissolving 10 g tryptone (Bacto, BD diagnostics), 10 g NaCl (Fisher Scientific), 5 g yeast extract (Bacto, BD diagnostics) in 1000 mL deionised H_2O (dH_2O) before autoclaving at 135°C, 1.6 bar for 25 min. LB agar contained an additional 15 g (1.5% w/v) agar (Bacto, BD diagnostics) prior to autoclaving. Once cooled to ambient temperature, supplements for antibiotic selection were added where required, and for liquid culture of *Mycobacterium smegmatis* 0.05% Tween80 (v/v final, Sigma) was added to prevent clumping (see section 2.1.5). LB broth was stored at room temperature and solid agar was stored at 4°C.

2.1.2 SOC broth

250 mL SOC (super optimal broth with catabolite repression) broth was made by dissolving 5 g tryptone, 1.25 g yeast extract with 500 µL 5 M NaCl, 625 µL 1 M KCl, 2.5 mL 1 M $MgCl_2$, 2.5 mL 1 M $MgSO_4$, 0.9 g glucose in 250 µL dH_2O prior to filter sterilisation. SOC broth was stored at -20°C in 10 mL aliquots prior to use.

2.1.3 7H9 broth

7H9 broth (Middlebrook) was made according to the manufacturer's specification, for 1 litre: 4.7 g 7H9 base and 4 mL glycerol (0.4% v/v final, Sigma) were dissolved in 900 mL dH_2O before autoclaving. Once cooled to ambient temperature 100 mL ADC (albumin dextrose catalase supplement, 10% v/v final, Middlebrook) was added with 0.05% (v/v final) Tween80 and supplements for antibiotic selection where required (see section 2.1.5). 7H9 broth was stored at room temperature.

2.1.4 7H11 agar

7H11 agar (Middlebrook) was made according to the manufacturer's specification, for 200 mL: 4.1 g 7H11 base and 1 mL glycerol (0.5% v/v final) were dissolved in 180 mL dH_2O before autoclaving. Once cooled to ambient temperature 20 mL OADC (oleic albumin dextrose catalase, 10% v/v final, Middlebrook) was added with supplements for antibiotic selection where required (see section 2.1.5). 7H11 solid agar was stored at 4°C.

2.1.5 Media supplements

The following supplements were added to media to enhance bacterial growth or facilitate selection (see Table 2.I).

Supplement	**Storage**	**[Stock] mg/mL**	**[Final] µg/mL**			
			Escherichia coli		**mycobacteria**	
			Liquid	**Solid**	**Liquid**	**Solid**
Ampicillin (MP Bio Science)	-20°C	100 in dH_2O	100	100	-	-
Hygromycin B (Cambridge Biosciences)	-20°C	50 in dH_2O	250	250	50	50
Kanamycin monosulfate (MP Bio Science)	-20°C	50 in dH_2O	50	50	50	20
X-gal (Sigma)	-20°C	50 in 100% DMSO	-	100	-	100
Tween80 (Sigma)	*4°C*	*20% (v/v in dH_2O)*	-	-	*0.05%*	-

Table 2.I media supplements

2.2 Bacterial culture

2.2.1 Culture strains

Transformed *Escherichia coli* were of a DH5α background (α-select bronze, Bioline). Transformed *Mycobacterium smegmatis* strains of a MC2_155 background (Snapper *et al.* 1990), *Mycobacterium bovis* Bacillus Calmette-Guérin (BCG) (Pasteur strain, ATCC 35734) and *Mycobacterium tuberculosis* of a H37Rv background (Cole *et al.* 1998).

2.2.2 Growth phase definitions

For all mycobacterial species cultured, exponential growth phase was harvested between an optical density (OD) of 0.5 and 0.8 as measured by light scattering at 600 nm ($OD_{600\,nm}$). Stationary growth phase was harvested 1 week after 1.0 $OD_{600\,nm}$ and late stationary growth phase was defined as growth >2 weeks after 1.0 $OD_{600\,nm}$ (Moores *et al.* 2017).

2.2.3 Growth conditions

M. tuberculosis was grown using 7H9 broth in polypropylene roller bottles (Cell Master, Griener Bio-One), rolling at 100 rpm. Time-course experiments were conducted as indicated.

Nutrient starvation was carried out by centrifuging exponential growth phase *M. tuberculosis* at 5000 g, 21°C for 10 min. Following centrifugation, pellets were re-suspended in 25 mL pre-warmed phosphate buffered saline (PBS) with 0.05% (v/v final) Tween80 then centrifuged for a further 10 min, repeating this process twice. Following the final wash, cells were re-suspended in PBS with 0.05% Tween80 up to the original volume of the culture.

Transcription inhibition of cultures was achieved with the addition of a final concentration of 200 µg/mL rifampicin (LKT Laboratories).

Re-growth from late stationary growth phase was carried out by diluting late stationary growth phase culture to 0.4 $OD_{600\,nm}$ with fresh pre-warmed 7H9 broth and incubated in a new roller bottle at 37°C.

2.2.4 *Mycobacterium bovis* BCG biofilm formation

M. bovis BCG was cultured on 7H11 agar supplemented with OADC. A single colony was picked using a 2 mm sterile loop and suspended in 10 mL 7H9 broth in a 50 mL centrifuge tube. Cultures were incubated static at 37°C for 14 days. Prior to inoculation media was checked for contamination by inoculating Columbia blood agar (Oxoid) and observing negative growth after 48 hours incubation at 37°C.

Inoculum was added to achieve a final inoculated culture density of 0.2 $OD_{600\ nm}$ in 10 mL 7H9 broth in 50 mL centrifuge tubes without the addition of Tween80, which would have disrupted biofilm aggregation. Cultures were inoculated in parallel with triplicate biological replicates for each time-point. Following inoculation, samples were sealed airtight to allow biofilm formation under non-aerated static culture at 37°C. Extracts for the 'week 0' time-point representative of planktonic growth were taken by centrifuging cultures at 5000 g, 4°C for 10 min. Following centrifugation supernatant was discarded and pellets were re-suspended in 1 mL 'RNApro' (MP Bio Science) in a 2 mL screw cap tube containing Lysing Matrix B (MP Bio Science) then stored at -20°C prior to sample extraction.

Biofilms were harvested at 3, 5, 8 and 12 weeks following inoculation. Pellicles were removed from the air-surface interface of cultures using a 10 mm sterile loop and suspended in 1 mL 'RNApro' (MP Bio Science) in a 2 mL screw cap tube containing Lysing Matrix B (MP Bio Science) by vigorous vortexing. Samples were stored at -20°C prior to extraction. RNA extraction of all samples was carried out in parallel as described below (see section 2.7.1).

2.3 Stock buffers

50x TAE buffer: 242 g Tris base (Trizma, Sigma) and 18.6 g Na_2-EDTA dissolved with 57.1 mL acetic acid in 1000 mL dH_2O.

10x TBE buffer: 108 g Tris base (Trizma, Sigma), 55 g boric acid and 9.3 g Na_2-EDTA in 1000 mL dH_2O.

10x PBS: 80 g NaCl, 2 g KCl, 14.4 g Na_2HPO_4 and 2.4 g KH_2PO_4 adjusted pH to 7.4 in 1000 mL dH_2O.

Z-buffer: 40 mM $NaH_2PO_4.H_2O$, 60 mM $Na_2HPO_4.7H_2O$, 1 M $MgSO_4.7H_2O$, 10 mM KCl in dH_2O (Miller 1972).

5x oligo annealing buffer: 10 mM Tris HCl pH 7.5 Sigma, 50 mM NaCl, 1 mM EDTA in dH_2O.

2x formamide stop buffer loading dye: 95% (v/v) formamide, 18 mM EDTA, 0.025% SDS (sodium dodecyl sulphate, w/v) and trace amounts of xylene cyanol and bromophenol blue in dH_2O.

2.4 DNA isolation and manipulation

2.4.1 DNA isolation and sequencing

2.4.1.1 Plasmid DNA isolation

All plasmid DNA isolations were carried out using the 'QIAprep Spin Miniprep Kit' (Qiagen) following the manufacturer's protocol. Overnight cultures of transformed *E. coli* were grown in 5 mL LB broth with suitable antibiotic selection in 20 mL universal tubes, incubated at 37°C, 200 rpm. Prior to extraction samples were centrifuged at 5000 g, 21°C for 3 min. After centrifugation the supernatant was decanted, re-suspending the cell pellet in 250 µL 'buffer P1' for kit extraction. DNA was eluted in 30 µL molecular grade H_2O to prevent complications during electroporation of mycobacteria. The concentration and purity of isolated DNA was recorded prior to storage at 4°C (see section 2.4.1.4).

2.4.1.2 Episomal DNA isolation

Episomal DNA extractions from transformed mycobacteria were carried out with a modified protocol using 'IntstaGene matrix' (Bio-Rad). A cell suspension was made by suspending an isolated colony in 1 mL sterile dH_2O or using 1 mL of dense culture grown from an isolated colony. Cell suspensions were centrifuged at 16000 g, 21°C for 1 min in 1.5 mL micro-centrifuge tubes. Following centrifugation, cell pellets were re-suspended in 200 µL of 'InstaGene matrix' (Bio-Rad) and incubated at 56°C for 15 to 30 min. After incubation the suspension was vortexed at high speed for 10 sec before incubation at 100°C for 8 min. After incubation, the suspension was vortexed again then centrifuged at 16000 g, 21°C for 3 min. After centrifugation, 18 µL of the resulting supernatant was used as template in a 50 µL PCR reaction (see section 2.4.6.1). The remaining extraction was stored at -20°C,

repeating the 100°C incubation and subsequent centrifugation if required for additional PCR.

2.4.1.3 Genomic DNA isolation

Genomic DNA (gDNA) was isolated from mycobacteria by harvesting culture grown on solid 7H11 agar and suspending in 300 µL TE buffer (Qiagen). Suspensions were heat-killed for 1 hour at 90°C, sealing tubes with plastic paraffin film (Parafilm M, Sigma). Following heat-killing, cells were centrifuged at 16000 g, 4°C for 5 min. Following centrifugation supernatant was discarded and cell pellets were stored at -20°C for subsequent use.

After thawing, heat-killed pellets were re-suspended in 300 mL TE buffer prior to addition of 15 µL fresh lysozyme/lipase solution to a final concentration of 20 mg/mL each and 2 µL of 10 mg/mL RNaseA (Sigma). Re-suspended pellets were then incubated at 37°C for 2 hours. Following incubation, cells were placed on ice to cool rapidly, then allowed to warm to room temperature. Once at room temperature, 7.5 µL of 20 µg/mL proteinase K (Sigma), 25 µL of 10% SDS (w/v) and was added to extracts and incubated at 50°C for 1 hour.

Following incubation extracts were purified by phenol-chloroform extraction, adding an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1, Ambion). Extracts were vortexed vigorously then centrifuged at 16000 g, 15°C for 15 min. Following centrifugation, the aqueous layer was transferred to a new 1.5 mL micro-centrifuge tube and mixed with 0.02 volumes 5 M NaCl, 0.1 volumes 3 M sodium acetate and 3 volumes 100% ethanol by gentle inversion. Extracts were precipitated at -20°C over 18 hours. Following precipitation, extracts were centrifuged at 16000 g, 4°C for 30 min. Following centrifugation the supernatant was discarded the pellet washed with 1000 µL 70% ethanol (v/v in DEPC treated H_2O) before centrifuging again at 16000 g, 4°C for 5 min. After centrifugation the supernatant was aspirated and the pellet allowed to air dry on ice for 5 min before re-suspending in DEPC treated H_2O to a concentration of 100 ng/µL. DNA concentration and purity was recorded prior to storing at -20°C (see section 2.4.1.4).

2.4.1.4 Nucleic acid quantification and sequencing

All nucleic acid concentrations were quantified using a spectrophotometer and associated software (Nanodrop 2000, Thermofisher Scientific). Nucleic acid purity was quantified by measuring the ratio of absorbance at 260-280 nm (1.80 for DNA and 2.00 for RNA). Protein contamination was recorded by measuring the ratio of absorbance at 230-260 nm. DNA was Sanger sequenced commercially by Source BioScience. Sequence data was analysed using CLC Main workbench (Version 7, Qiagen).

2.4.2 DNA restriction digestion

All DNA restriction digests were carried out using New England Biolabs restriction enzymes. Digests were carried out in 0.5 mL micro-centrifuge tubes. For cloning vector digests 750 ng plasmid was added per reaction, for purified PCR fragment digestion or fragment generation from plasmidic DNA up to 10 µg plasmidic DNA was added. DNA was mixed with 1 µL restriction enzyme (1 µL of each enzyme for double digests) and 5 µL 10x 'CutSmart buffer' (NEB) in 50 µL with molecular grade H_2O. Reactions were incubated at 37°C for 18 hours before heat inactivation at 80°C for 5 min. Digested DNA was stored at 4°C prior to use. Cloning vector digests proceeded directly to dephosphorylation (see section 2.4.3).

2.4.3 DNA dephosphorylation

Dephosphorylation of restriction digested DNA was carried out using 'Antarctic phosphatase' (NEB). In a 0.5 mL micro-centrifuge tube digested DNA was mixed with 1 µL enzyme and 5 µL 10x reaction buffer (NEB) in 50 µL with molecular grade H_2O. Reactions were incubated at 37°C for 1 hour before heat inactivation at 70°C for 5 min. Dephosphorylated DNA was stored at 4°C prior to use.

2.4.4 Oligonucleotide annealing and phosphorylation

Complementary oligonucleotides were designed to form restriction digest compatible ends once annealed. Each reconstituted oligonucleotide was 5'phosphorylated using T4 polynucleotide kinase (NEB) separately in a 1.5 mL micro-centrifuge tube by combining 30 µL 100 µM oligonucleotide, 5 µL 10x reaction buffer (NEB), 5 µL 10 mM ATP (NEB) and 1 µL T4 PNK enzyme in 50 µL with molecular grade H_2O. Reactions were incubated at 37°C for 45 min before heat inactivation at 65°C for 10 min. In a 1.5 mL micro-centrifuge tube

22.5 µL of each complementary phosphorylated oligonucleotide were added to 5 µL 10x annealing buffer (see section 2.3) then incubated at 95°C for 5 min before being left to cool to room temperature over >1 hour. Annealed oligonucleotides were stored at 4°C prior to ligation with compatible restriction digested DNA.

2.4.5 DNA ligation

All DNA ligations were carried out using 'T4 DNA ligase' (Promega) according to the manufacturer's specification. Ligations were carried out in a 0.5 mL micro-centrifuge tube as a 10 µL reaction composed of 7 µL digested DNA insert, 1 µL digested and dephosphorylated compatible DNA vector, 1 µL 10x 'DNA ligase buffer' (Promega) and 1 µL T4 DNA ligase. Reactions were incubated at 16°C for 18 hours before being transformed into competent *E. coli*.

For pGEM (Promega) T-A ligations, rapid ligase buffer was used. Ligations were carried out as a 10 µL reaction composed of 3 µL 3'terminal deoxyadenylated PCR product (see section 2.4.6.2), 1 µL pGEM vector, 5 µL 2x Rapid Ligation buffer (Promega) and 1 µL T4 DNA ligase. Reactions were incubated at 16°C for 1 hour before being transformed into competent *E. coli*.

2.4.6 Polymerase chain reaction DNA amplification

Oligonucleotides were designed to flank desired target sequences. If the desired amplicon was to be used for cloning, oligonucleotides incorporated restriction digestion sites with flanking nucleotides at the 5'end to facilitate digestion. PCR reactions were carried out in a thermocycler with annealing temperatures corresponding to 3°C below primer melting temperatures (T_m) as predicted by Sigma. For primer pairs with disparate T_m a 'touch down' PCR was carried, lowering the annealing temperature from up to 10°C above the lower T_m by -1°C per cycle. Extension time was set according the length of the DNA fragment to be amplified. Following the completed cycles, PCR reactions were stored at 4°C and analysed by agarose gel electrophoresis. 5% DMSO (dimethyl sulfoxide, v/v final, Sigma) was used in PCR reactions to facilitate DNA melting and improve amplification.

2.4.6.1 Phusion GC high fidelity PCR

'Phusion GC high fidelity DNA polymerase master mix' (NEB) was used to amplify desired DNA sequences with high fidelity to generate amplicons for cloning or DNA sequencing. Each reaction was carried out in a 0.25 µL PCR tube with 12.5 µL enzyme mix (NEB), 1 µL 10 µM forward primer, 1 µL 10 µM reverse primer, 2.5 µL 50% DMSO, 1 µL 10 ng/µL template plasmid in 25 µL with molecular grade H_2O. Reactions were amplified over 30 repeat cycles (see Table 2.II)

Step	Temperature	Time	Number of cycles
Initial denaturation	98°C	10 min	1
Denaturation	98°C	30 sec	
Annealing	72 – 50°C	45 sec	30
Extension	72°C	1 min/kb	
Final extension	72°C	10 min	1

Table 2.II example Phusion GC PCR thermocycling conditions

Fragments for cloning were purified using the 'QIAquick PCR Purification Kit' (Qiagen) following the manufacturer's protocol and eluting in molecular grade H_2O. Fragments for T-A ligation into pGEM included an additional 2 cycle REDTaq PCR step for 3'deoxyadenylated end generation (see section 2.4.6.2).

2.4.6.2 REDTaq PCR

'REDTaq ReadyMix' (Sigma) was used for colony PCR and the production of 3'deoxyadenylated PCR products. Reactions were carried out in a 0.25 µL PCR tube with 12.5 µL enzyme mix (Sigma), 1 µL 10 µM forward primer, 1 µL 10 µM reverse primer, 2.5 µL 50% DMSO, 1 µL 10 ng/µL template plasmid in 25 µL with molecular grade H_2O. Bacteria were added directly to colony PCR reactions by suspending an isolated colony in the PCR reaction, releasing DNA template through cell lysis during the initial denaturation. Reactions were amplified over 30 repeat cycles as in Table 2.II but with denaturation at 94°C.

For 3'deoxyadenylated PCR products for pGEM T-A ligation, after the initial 25 µL Phusion GC high fidelity PCR reaction was complete, 25 µL 'REDTaq ReadyMix' PCR reaction mix was added to reactions for a further 2 cycles with 94°C denaturation.

2.4.7 PCR mutagenesis

2.4.7.1 Q5 site directed mutagenesis

Site-directed mutagenesis (SDM) was carried out using the 'Q5 SDM kit' (NEB) following the manufacturer's protocol. Appropriate primers were designed using the 'NEBasechanger' SDM primer design tool with additional flanking regions to optimise for high G:C content. Potential clones were screened by colony PCR, plasmid DNA isolation and sequencing. Correct constructs were sub-cloned into an un-mutated vector backbone to prevent undesired mutations, confirming construction by colony PCR, plasmid DNA isolation and sequencing.

2.4.7.2 Overlap extension mutagenesis

For small mutations a pair of Phusion GC polymerase PCR reactions were carried out: reaction (A) used an upstream forward primer and a mutagenic reverse primer spanning the region to be mutated, reaction (B) used a downstream reverse primer and a mutagenic forward primer spanning the region to be mutated. The resulting amplicons contained a region of complementarity exploited in reaction (C) by combining 1 µL of each as template in another PCR reaction with the non-mutagenic primers of the original reactions. Desired reaction (C) amplification was confirmed by agarose gel electrophoresis, purifying the remaining reaction using the 'QIAquick PCR Purification Kit' (Qiagen) following the manufacturer's protocol and eluting in molecular grade H_2O. After purification, DNA was digested with restriction enzymes, ligated into the desired vector and transformed into *E. coli*. Desired mutations were confirmed by colony PCR, plasmid DNA isolation and sequencing.

2.4.8 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate nucleic acids by size. All gels were run at 120 V constant to achieve separation. Gels were analysed using 'GeneFlash Bio Imaging Gel Documentation System' (Syngene). Gel images were obtained using 'Typhoon FLA 9500' (GE). Long DNA oligonucleotides (>1000 bp) were separated on 1% agarose (w/v, MP Bio Science) in 1x TAE buffer (see section 2.3) stained with ethidium bromide (Sigma), estimating size with 5 µL 'BenchTop 1 kb DNA Ladder' (Promega). Shorter DNA

oligonucleotides (<1000 bp) were separated on 2% (w/v) agarose estimating size with 5 µL 'BenchTop 100 bp DNA Ladder' (Promega). For PCR amplicon analysis, 5 µL of DNA was mixed with 1 µL 6x 'blue/orange loading dye' (Promega) prior to loading. RNA was analysed using 1-3% 'QA-Agarose High Resolution' (w/v, MP Bio Science) in 1x TAE buffer stained with ethidium bromide; RNA was diluted in up to 5 µL DEPC treated H_2O (Ambion) and mixed with 1 µL 6x 'blue/orange loading dye' prior to loading.

2.4.8.1 Gel extraction of DNA

For gel extraction purification, 45 µL DNA solution was mixed with 5 µL 6x 'blue/orange loading dye' prior to 1% agarose gel electrophoresis. After electrophoresis the desired amplicon was excised under UV light and extracted using the 'QIAquick Gel Extraction Kit' (Qiagen) following the manufacturer's protocol and eluting in molecular grade H_2O.

2.5 Transformation of DNA

2.5.1 Transformation of *Escherichia coli*

Commercially produced competent cells (α-select bronze, Bioline) were used for all *E. coli* transformations. In a 1.5 mL micro-centrifuge tube 50 µL of cells were mixed with either 1 µL (>50 ng/µL) purified plasmid or 10 µL ligation reaction and incubated on ice for 20 min. Following incubation the cells were heat shocked at 42°C for 1 min before immediately chilling on ice for 1 min. Once cooled, 900 µL of SOC media was added for recovery at 37°C, 200 rpm for 1 hour. Following incubation cells were centrifuged at 16000 g, 21°C for 1 min, 700 µL supernatant was aspirated and the remaining 200 µL used for cell pellet re-suspension and inoculation of LB agar plates with appropriate selection. Plates were incubated at 37°C overnight.

2.5.2 Protocol for competent mycobacteria

Competent mycobacteria were made using exponential growth phase culture. *M. smegmatis* was grown in LB 0.05% (v/v final) Tween80 at 37°C, 200 rpm. Culture was inoculated with a dense starter culture for growth to exponential phase over >6 generations. At 0.8 $OD_{600\ nm}$ 50 mL culture was decanted into 50 mL centrifuge tubes and centrifuged at 5000 g, 4°C for 10 min. Following centrifugation supernatant was decanted

and the cell pellet re-suspend in decreasing volumes of chilled 10% glycerol (v/v), centrifuging in-between at 5000 g, 4°C for 10 min. Re-suspension volume was halved each time from 25 to 12.5 and finally 5 mL (10% of the original volume of culture). The final suspension was aliquoted and stored at -20°C. The same process was used for *M. tuberculosis* using 7H9 0.05% (v/v final) Tween80 instead of LB broth.

2.5.3 Transformation of competent mycobacteria

The electroporation protocol was adapted from Wards and Collins (1996). Electroporation cuvettes (0.2 cm Gene Pulser/MicroPulser, Bio-Rad) were stored at -20°C prior to use. 400 µL competent mycobacteria were added to the electroporation cuvette along with 4 µL (>50 ng/µL) construct DNA, and then incubated on ice for 10 min. Following incubation, cuvettes were gently agitated to disperse cells then subjected to a single pulse of 2.5 kV, 25 µF and 1000 Ω in the electroporator (Gene Pulser Xcell, Bio-Rad) incubating on ice immediately after for 10 min.

After being chilled, electroporated *M. smegmatis* were suspended in 3.6 mL LB 0.05% (v/v final) Tween80 for recovery at 37°C, 200 rpm for 3 hours. After incubation the cells were centrifuged at 5000 g, 21°C for 10 min, decanting the supernatant to leave ~200 µL for re-suspension and inoculation of LB agar with suitable selection. Plates were incubated at 37°C for 1 week. The same process was used for *M. tuberculosis* using 7H9 0.05% (v/v final) Tween80 instead of LB broth. As a slower growth strain, *M. tuberculosis* was recovered over 24 hours, at 37°C, 100 rpm. Cells were plated onto 7H11 agar instead of LB and incubated at 37°C for 2 weeks.

2.6 β-galactosidase expression assay

2.6.1 Protein extraction

Total protein extracts were obtained from *M. smegmatis* cultures grown in LB 0.05% (v/v final) Tween80 with suitable selection. At 0.7 $OD_{600\ nm}$ 30 mL of culture was transferred to 50 mL centrifuge tubes and incubated on ice for 10 min to prevent centrifugation stress. Following incubation cells were centrifuged at 5000 g, 4°C for 10 min. After centrifugation cells were re-suspended in 5 mL Z-buffer (see section 2.3) gently vortexed then centrifuged

again at 5000 g, 4°C for 10 min. Following centrifugation cells were re-suspended in 1 mL Z-buffer and transferred to 1.5 mL micro-centrifuge tubes for centrifugation at 16000 g, 4°C for 10 min. Cells were re-suspended and centrifuged in 1 mL Z-buffer a further two times before being transferred to 2 mL screw cap tubes containing ~0.2 mL 150-212 µm glass beads (Sigma) for lysis in a sample homogeniser (FastPrep-24 5G, MP Bio Science) at 6.5 m/s for 30 sec. Following lysis and centrifugation at 16000 g, 4°C for 10 min, supernatant was aliquoted into 1.5 mL micro-centrifuge tubes and stored at 4°C prior to use.

2.6.2 β-galactosidase expression quantification

In a 1.5 mL micro-centrifuge tube 38 mM final β-mercaptoethanol (Sigma) was added to 500 µL total protein extract to enhance β-galactosidase activity along with 100 µL 4 mg/mL OPNG substrate (ortho-nitrophenyl-β-galactoside, Sigma) prior to incubation at 28°C. Upon production of strong yellow colour, reactions were stopped by adding 250 µL 1 M $NaCO_3$ and the in incubation time noted. 200 µL of each sample was then measured in triplicate in a 96 well plate for 420 nm absorbance in a microplate reader (FLUOstar OPTIMA, BMG Labtech).

2.6.3 BCA protein concentration quantification

Protein concentration quantification of whole protein extracts was carried out using by 'Pierce BCA Protein Assay Kit' (Thermo Scientific) following the manufacturer's protocol. 25 µL of each sample was tested in triplicate in a 96 well microplate with a set of standards, measuring 562 nm absorbance in a microplate reader to calculate the total protein concentration against a standard curve.

2.6.4 β-galactosidase assay statistics

Quantification of β-galactosidase expression was calculated in Miller units:

$$\frac{A_{420}\times 380}{\min\times\, mg\ total\ protein} = \frac{Miller\ units}{mg\ protein}$$

Miller units were expressed as a percentage of the average WT value.

Statistical significance was calculated using one-way ANOVA with Tukey post hoc analysis in IBM SPSS ±1 standard deviation. Significance thresholds specified as: 'NS' (no significant difference, $P > 0.05$), '*' ($P \leq 0.05$), '**' ($P \leq 0.01$) and '***' ($P \leq 0.001$).

2.7 Extraction and manipulation of RNA

2.7.1 RNA extraction

Mycobacterial RNA extraction was carried out using 'FastRNA Pro Blue Kit' (MP Bio Science) following the manufacturer's protocol. Roughly 15 mL of chipped ice was added to 50 mL centrifuge tubes and pre-chilled at -20°C. At the time of extraction, 25 mL of bacterial culture was added to the 50 mL pre-chilled tube on ice, quickly proceeding to centrifugation at 5000 g, 4°C for 10 min. After centrifugation, supernatant was decanted and the pellet (or pellets of larger extraction volumes) re-suspended in 'RNApro' (MP Bio Science), 1 mL per 25 mL original culture and transferred to kit Lysing Matrix B 2 mL screw cap tubes. The sample was then processed using a sample homogeniser (FastPrep-24 5G, MP Bio Science) at 6.0 m/s for 40 sec.

After homogenisation samples were immediately returned to ice then centrifuged at 16000 g, 4°C for 5 min, transferring the upper phase to a 1.5 mL micro-centrifuge tube and incubating at room temperature for 5 min. After incubation, 300 µL chloroform (Sigma) was added; samples were vortexed vigorously then incubated again at room temperature for 5 min. After incubation samples were centrifuged at 16000 g, 4°C for 5 min, transferring aqueous phase to a 1.5 mL micro-centrifuge tube with 500 µL 100% ethanol for precipitation at -20°C for 18 hours.

After precipitation, samples were centrifuged at 16000 g, 4°C for 30 min, decanting the supernatant and washing the RNA pellet with 1000 µL 70% ethanol (v/v in DEPC treated H_2O) before centrifuging again at 16000 g, 4°C for 5 min. After centrifugation the supernatant was aspirated and the pellet allowed to air dry on ice for 5 min before re-suspending in 25 µL DEPC treated H_2O. RNA concentration and purity was recorded prior to storing at -80°C (see section 2.4.1.4). RNA integrity was analysed by agarose gel electrophoresis of 1000 ng extracted RNA.

2.7.2 DNase treatment of total RNA

DNase treatment was carried out using 'TURBO DNase I' (Ambion) adapting the manufacturer's protocol. In a 1.5 mL micro-centrifuge tube 10 µg total RNA was combined with 5 µL 10x reaction buffer (Ambion), 1 µL RNase inhibitor (RNasin Plus, Promega) and 2 µL enzyme in 50 µL with DEPC treated H_2O. Reactions were incubated at 37°C for 30 min. After 30 min another 2 µL enzyme was added, incubating reactions for a further 30 min. After incubation reactions were stopped immediately by proceeding with a phenol-chloroform extraction adapted from Gehart, Wagner and Vogel (2008). For each 50 µL treated RNA, 150 µL DEPC treated H_2O and 200 µL pH 4.5 phenol-chloroform-isoamyl alcohol (125:24:1, Ambion) were added, vortexed vigorously then centrifuged at 16000 g, 15°C for 15 min. After centrifugation, the aqueous layer was transferred to a new 1.5 mL micro-centrifuge tube and mixed with 600 µL 100% ethanol, 6 µL 3 M sodium acetate and 2 µL glycogen co-precipitant (GlycoBlue 15 mg/mL, Ambion) for precipitation at -20°C for 18 hours. If DNased RNA was to be used directly for cDNA synthesis, 1 µL 50 ng/µL random hexamers (Promega) were added during precipitation.

After precipitation, samples were centrifuged at 16000 g, 4°C for 30 min, decanting the supernatant and washing the RNA pellet with 1000 µL 70% ethanol (v/v in DEPC treated H_2O) before centrifuging again at 16000 g, 4°C for 5 min. After centrifugation the supernatant was aspirated and the pellet allowed to air dry on ice for 5 min. RNA pellets were re-suspended in an appropriate volume of DEPC treated H_2O for downstream processes and stored at -80°C prior to use. 1 µL re-suspended RNA was used as template in a REDTaq PCR reaction primed with oligonucleotides 5.51 and 5.52 (see Appendix section 10.3) to check for genomic DNA from any mycobacterial species by agarose gel electrophoresis. DNase treatment was repeated as necessary until no genomic DNA was detectable.

2.7.3 Rapid amplification of cDNA ends

The rapid amplification of cDNA ends (RACE) protocol was adapted from Gehart, Wagner and Vogel (2008) and GeneRacer kit (Invitrogen).

2.7.3.1 3'polyadenylation

All 3'polyadenylations were carried out using a 'Poly(A) Tailing Kit' (Ambion) modifying the manufacturer's protocol. In a 1.5 mL micro-centrifuge tube 10 µg DNase treated RNA was re-suspended in 15 µL DEPC treated H_2O was combined with kit reagents: 5 µL 'E-PAP buffer', 2 µL 25 mM $MnCl_2$, 2 µL 10 mM ATP and 1 µL *E. coli* poly(A) polymerase (E-PAP). Reactions were incubated at 37°C for 30 min before proceeding directly with phenol-chloroform extraction. Extraction was carried out as described previously (see section 2.7.2) but with the addition of 1 µL GeneRacer oligo d(T) primer (oligonucleotide 2.07, see Appendix section 10.3) during precipitation that included a 3'adaptor sequence. Extracted 3'polyadenylated RNA was re-suspended in 12 µL DEPC treated H_2O for cDNA synthesis (see section 2.7.4) and stored at -80°C prior to use.

2.7.3.2 Construction of RACE library

RACE treated cDNA was amplified in a 50 µL REDTaq polymerase reaction over 40 cycles using an adaptor specific and gene specific primer pair. Reactions for 3'RACE used an upstream gene specific primer (oligonucleotide 2.15) and adaptor specific primer (oligonucleotide 2.09, see Appendix section 10.3), which complemented the 3'adaptor sequence. Gene specific amplification was confirmed by agarose gel electrophoresis of 5 µL PCR product on 3% agarose. The remaining PCR product was used for RACE library construction directly.

All RACE amplicons were ligated into PCR4 TOPO cloning vector using 'TOPO TA Cloning Kit' (Invitrogen). In a 1.5 mL micro-centrifuge tube 4 µL RACE amplicon was mixed with kit reagents: 0.5 µL salt solution (1.2 M NaCl, 0.06 M $MgCl_2$) and 0.5 µL vector (10 ng plasmid in 50% glycerol, 50 mM Tris-HCl pH 7.4, 1 mM EDTA, 2 mM DTT, 0.1% Triton X-100, 100 µg/mL BSA and 30 µM phenol red). Reactions were incubated at room temperature for 5 min then transformed into *E. coli* as previously described, inoculating LB agar with kanamycin selection for incubation at 37°C for 18 hours. Individual colonies were isolated for plasmid

isolation and sequencing using M13 forward and reverse primers (oligonucleotides 2.38 and 2.39, see Appendix section 10.3) to generate a consensus sequence for the target RNA terminus. The first non-adenosine residue beyond the 5'end of the 3'polyadenlyated tail was regarded as the 3'terminal residue of the transcript. Where the 3'polyadenlyated tail of reads mapped to a potential adenosine residue the nearest upstream non-adenosine residue was regarded as the 3'terminal residue of the transcript to prevent ambiguity.

2.7.4 cDNA synthesis

All reverse transcription PCRs were carried using out using 'SuperScript III Reverse Transcriptase' (Invitrogen). The protocol was adapted using the SuperScript and GeneRacer kit (Invitrogen) protocols as well as Gehart, Wagner and Vogel (2008). In a 1.5 mL micro-centrifuge tube 10 µg DNase treated RNA was re-suspended in 12 µL DEPC treated H_2O. For reverse transcriptase analysis 10 µg treated RNA was re-suspended in 24 µL, and split into duplicate 12 µL aliquots for reverse transcriptase (RT) +/- reactions. All RNA pellets contained primer added during purification steps.

RNA was mixed with 1 µL dNTP mix (10 mM of each dNTP in DEPC treated H_2O) then denatured at 65°C for 5 min before being stored on ice. After denaturing the RNA suspension was transferred to 0.25 mL PCR tubes and mixed with kit components: 4 µL 5x reaction buffer, 1 µL 0.1 M DTT, 1 µL RNase inhibitor (RNasin Plus, Promega) and 1 µL enzyme (or 1 µL DEPC treated H_2O for RT- reactions). Reactions were incubated in a thermocycler at 50°C for 30 min, 55°C for 30 min, heat inactivated at 70°C for 15 min before being immediately stored on ice. Once cooled, 1 µL RNase H (NEB) was added to RT+ reactions only and then all reactions were incubated at 37°C for 20 min. After incubation samples were stored at -20°C prior to use.

2.7.5 Analysis of RT-PCR

Co-transcription of *rpfB* and *ksgA* was analysed using 2 µL of RT+/- cDNA generated with random hexamers (Promega). cDNA was amplified in a 50 µL REDTaq PCR reaction over 30 cycles primed using oligonucleotides 5.51 and 8.04 (see Appendix section 10.3) which flanked the *rpfB-ksgA* overlapping region complementary to the sequence in

M. tuberculosis, M. bovis BCG and *M. smegmatis*. Amplification was confirmed by agarose gel electrophoresis. Gel images were obtained using 'Typhoon FLA 9500' (GE).

2.7.6 Real-time quantitative PCR

'SensiFast SYBR Hi-ROX master mix' (Bio-line) was used to amplify cDNA for real-time quantitative PCR (RT-qPCR). A master mix consisting of 5 µL enzyme mix, 1.5 µL DEPC treated H_2O, 0.25 µL of 10 µM forward primer and 0.25 µL of 10 µM reverse primer was distributed into 7 µL aliquots in a 96 well plate (MicroAmp Fast Optical 96-Well Reaction Plate, Applied Biosystems). *M. tuberculosis* H37Rv gDNA was used to create a standard curve for each reaction by serially diluting from 1 ng/µL to 0.00001 ng/µL in DEPC treated H_2O. Wells were loaded with either 1 µL standard in three technical replicates or 1 µL RT+/- cDNA in 4 technical replicates. Plates were then sealed with an adhesive foil (MicroAmp Optical Adhesive Covers, Applied Biosystems) and stored at 4°C prior to use.

Plates were centrifuged at 1000 g for 1 min prior to running. All reactions were carried out using a 'QuantiStudio 6 Flex Real-Time PCR System' (Applied Biosciences) and analysed using QuantiStudio Real-time PCR software v1.1 (Applied Biosciences). Reactions were amplified over 40 cycles with a high-resolution melt to confirm amplicon size (see Table 2.III) Quantification was calculated according to the cycle at which amplification exceeded an arbitrary threshold (typically cycle 13 to 15) set by amplification of a *M. tuberculosis* gDNA standard (serial dilution from 1 to 0.00001 ng/µL), expressed as arbitrary units calculated against the standard curve.

Step	Temperature	Rate	Time	Number of cycles
Initial denaturation	95°C	2.42°C/s	3 min	1
Denaturation	95°C	2.42°C/s	5 sec	
Annealing	60°C	2.42°C/s	10 sec	40
Extension	72°C	2.42°C/s	10 sec	
Final denaturation	95°C	2.42°C/s	15 sec	1
Annealing	60°C	2.42°C/s	1 min	1
High resolution melt	From 60 to 95°C	0.05°C/s	-	-

Table 2.III example RT-qPCR thermocycling conditions

None of the cDNA analysed by RT-qPCR was generated in a strand specific manner, because of this it was necessary to prime different regions of the *rpfB* locus select desired targets

specifically (see Table 2.IV). As riboswitch amplicons might originate from attenuated transcripts or read-through, the riboswitch was primed forward upstream of the P_2 TSS and reverse within the *rpfB* coding sequence (CDS) to obtained a value of riboswitch read-through into the CDS. The fraction of riboswitch read-through was derived by dividing read-through values over riboswitch only values (see Figure 2.I), and expressed as a percentage.

Amplicon	Target	Forward primer	Reverse primer
① **107 bp**	*asRNA*	8.58, upstream of P_1 TSS	8.53, within *riboswitch*
② **72 bp**	*riboswitch*	8.11, upstream of P_2 TSS	8.53, within *riboswitch*
③ **332 bp**	P_1 *rpfB* read-through	8.56, upstream of P_2 TSS	8.57, within *rpfB* CDS
④ **235 bp**	*rpfB* from P_1 or P_2	8.07, within *rpfB* CDS	8.08, within *rpfB* CDS
⑤ **231 bp**	*ksgA* from P_1 or P_2	8.09, within *ksgA* CDS	8.10, within *ksgA* CDS

Table 2.IV target amplicons of the *rpfB* locus
Numbers in circles correlate to amplicons depicted in Figure 2.I. Primer oligonucleotides are detailed in Appendix section 10.3.

③ / ② = P_1 *rpfB* read-through over riboswitch

(③ / ②) x 100 = P_1 *rpfB* read-through as percentage riboswitch

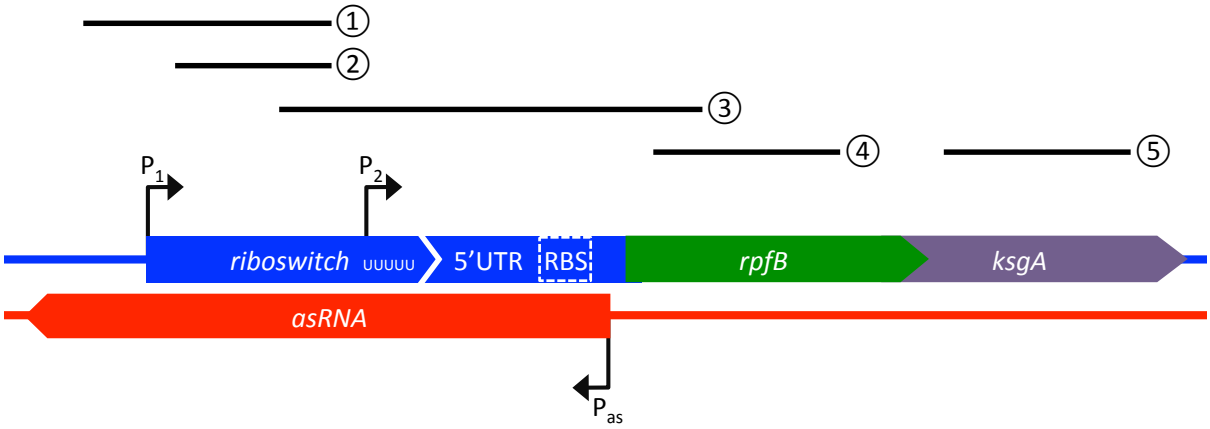


Figure 2.I schematic of RT-qPCR target amplification
Diagram depicts the target amplicons within the *rpfB* locus. Circled numbers correlate with amplicons described in (see Table 2.IV).

2.8 Northern blotting of RNA

2.8.1 Preparation of RNA for gel electrophoresis

Total RNA sample concentration was standardised at either 0.5 or 1 µg/µL by suspending the required mass of RNA in 200 µL with DEPC treated H_2O (Ambion) in a 1.5 mL micro-centrifuge tube precipitating with 600 µL 100% ethanol at -20°C overnight. Samples were then centrifuged at 16000 g, 4°C for 30 min, decanting the supernatant and washing the pellet with chilled 70% ethanol (v/v in DEPC treated H_2O, Ambion) before re-suspending the RNA pellet in loading dye (Gel Loading Buffer II, Ambion) to achieve the desired concentration. Before loading the gel, the required volume of RNA was aliquoted into a 1.5 mL micro-centrifuge tube, denatured at 80°C for 2 min then stored on ice. Samples were loaded alongside 10 µL 0.5 µg/µL RNA marker (Century, Ambion) suspended in loading dye to estimate RNA size.

2.8.2 Urea polyacrylamide gel electrophoresis

RNA was separated using denaturing urea denaturing urea polyacrylamide gel electrophoresis (PAGE): 7 M urea, 1x TBE buffer (see section 2.3) and 8-15% (v/v final) of 40% 19:1 acrylamide:bis-acrylamide (MP Bio Science) in 50 mL with dH_2O, cross-linked with 100 µL of 25% APS (ammonium persulfate, w/v in dH_2O) and 25 µL TEMED (N,N,N',N-tetramethylethylenediamine), stored at 4°C prior to use. Gels were pre-run at 15 W constant for 1 hour to warm the gel prior to loading. Once loaded, gels were run at 12 W constant for 15 min and then reduced to from 11 to 9 W constant, using loading dye to estimate RNA migration.

2.8.3 Northern electro-blotting

Migrated gels were electro-blotted onto nylon membrane (BrightStar-Plus Positively Charged Nylon Membrane, Ambion) between six layers of Whatman 3MM paper (Sigma) wet with 0.5x TBE buffer (see Figure 2.II).

Electro-blotting was carried out at 7 V constant for 1 hour. Once transferred the blotted membrane was cross-linked on both sides (120,000 µJ, UV Stratalinker 2400, Stratagene) then stained for 10 min in a bath of 0.3 M sodium acetate with 0.03% (w/v) methylene blue.

Scoring the membrane with pencil before de-staining with dH_2O enhanced the RNA marker. Membranes were in Whatman 3MM paper at -20°C prior to hybridisation.

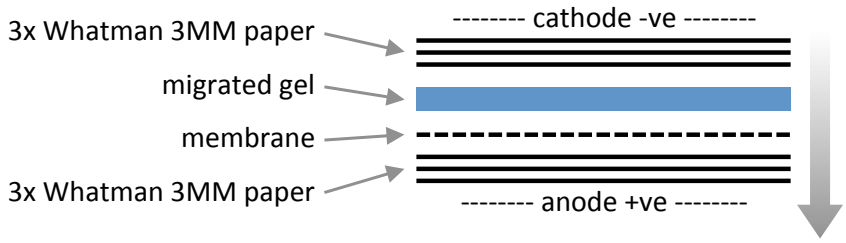


Figure 2.II construction of the northern electro-blot

2.8.4 Ribo-probe construction

Ribo-probes were constructed using the 'mirVana miRNA Probe Construction Kit' (Ambion) following the manufacturer's protocol. Template oligonucleotide was commercially synthesised by Sigma with 40 nt of target RNA complementary sequence with 3'incorporated T7 promoter primer complementary sequence 'GCCTGTCTC' (see Appendix section 10.3). Template oligonucleotide was used to generate dsDNA template for radiolabelled ribo-probes construction using the kit, incorporating 32P-αUTP (3000 Ci/mmol 10 mCi/mL, Perkin Elmer) by adding 5 µL to the reaction (doubling volume with each half-life) and incubating with T7 RNAP for 10 min. The DNase I treatment step was not observed. Synthesised radiolabelled ribo-probes were added to pre-hybridised membranes immediately, with the exception of 5S rRNA probe which was stored at -20°C using 2 µL at a time due to the abundance of 5S rRNA transcripts.

2.8.5 Ribo-probe hybridisation

Blotted membranes were pre-hybridised with approximately 10 mL hybridisation buffer (ULTRAhyb Ultrasensitive Hybridization Buffer, Ambion), in hybridisation bottles at rolling at 68°C for 1 hour. Once pre-hybridised the freshly transcribed ribo-probe was added directly into the hybridisation buffer for hybridisation at 68°C for 18 hours. Following probe hybridisation the buffer was discarded and the membrane successively washed with once with 2x SSC (salt sodium citrate buffer, 20x stock, MP Bioscience) with 1% SDS (w/v) and twice with 0.2x SSC 1% SDS, incubating each wash rolling at 68°C for at least 30 min and decanting previous washes. Finally the membrane was washed briefly with 2x SSC before

being dried and wrapped in Saran wrap to expose to a phosphor screen developed using a Typhoon FLA 9500 (GE).

2.9 *Escherichia coli* RNAP *in vitro* transcription system

2.9.1 *In vitro* transcription reaction

E. coli RNAP *in vitro* transcription assays were carried out using previously described methods for producing halted transcription elongation complexes (TECs) (Landick, Wang & Chan 1996; Grundy, Winkler & Henkin 2002; Sudarsan *et al.* 2003; Wickiser *et al.* 2005). Transcription templates were cloned into pGAMrnnX, a template vector derived from pGEM (Promega) containing the *M. smegmatis rrnB* P_1 promoter region with extended -10 element (see Appendix section 10.2). Transcription templates were comprised of -80 to -1 of the P_{rrnB} TSS, followed downstream with the sequence 'GTAAA' fused with the *rpfB* riboswitch sequence spanning from the P_1 TSS +1 to +176 bp to generate riboswitch-mediated attenuation at ~130 nt and transcription read-through as template run-off at 181 nt or synthetic termination by the *synB* terminator at 211 nt (see section 7.2).

Templates were generated by PCR amplification with oligonucleotides 5.44 and 5.61/8.43 (see Appendix section 10.3) with Phusion GC polymerase. Amplified template was purified by ethanol precipitation in 2.5 volumes 100% ethanol and 0.1 volumes 3 M sodium acetate, precipitating at -20°C for 18 hours. After precipitation templates were centrifuged at 16000 g, 4°C for 30 min, decanting the supernatant and washing the pellet with 1000 µL 70% ethanol (v/v in DEPC treated H_2O) before centrifuging again at 16000 g, 4°C for 5 min. After centrifugation the supernatant was aspirated and the pellet allowed to air dry on ice for 5 min before re-suspending in 50 µL DEPC treated H_2O. DNA concentration and purity was recorded prior to storing at -20°C (see section 2.4.1.4). All *rpfB* riboswitch templates were cloned with an additional 5'sequence of 'GTAAA' to facilitate transcriptional pausing. The dinucleotide GpU (IBA) was used to initiate transcription whilst withholding rUTP, this caused TECs to halt at the first uracil at U+11 nt (with the exception of the anti-terminator stabilised mutant U11C, which paused at U+31 nt).

All *E. coli* RNAP transcription reactions were carried out using commercially produced core or holoenzyme RNAP (NEB). Halted TECs were assembled first with the addition of 5x

reaction buffer (40 mM Tris-HCl pH 7.5, 150 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.01% Triton X-100), 20 nM dsDNA template, 40 U 'RNasin plus' RNAse inhibitor (Promega) and 34 nM *E. coli* holoenzyme in 22.5 µL with DEPC treated H_2O. Enzyme and template were incubated at 37°C for 5 min before adding 5 µL pre-warmed 10x halt mix (150 µM GpU dinucleotide, 25 µM rGTP/rCTP with 10 µM rATP (Promega) and 0.04 µCi/µL 32P-αATP (3000 Ci/mmol 10 mCi/mL, Perkin Elmer). Reactions were then incubated at 37°C for 5 min. After incubation pre-warmed heparin (Sigma) was added to 200 ng/µL final in up to 65 µL with 1x buffer to block additional transcription initiation, halted TEC reactions were then stored on ice. Ligand was added to reactions in 0.25 mL tubes by adding 65 µL halted TEC to with ligand up to 90 µL in 1x buffer. Reactions were incubated at 37°C for 5 min then transcription elongation activated with the addition of 10 µL 10x elongation mix (250 µM of all rNTPs mix) at 25 µM final in 100 µL. Reactions were incubated at 37°C for 5 min before being stopped with 100 µL 2x formamide stop buffer loading dye (see section 2.3). Reactions were stored at -20°C prior to use.

2.9.2 *E. coli* biotin-streptavidin roadblock assay

2.9.2.1 Biotinylated template coupling

Biotinylated templates were generated by PCR amplification as above using oligonucleotides 5.44 and 8.55 (see Appendix section 10.3) using Phusion GC polymerase. The reverse primer incorporated a 5'biotin tag onto the antisense strand of the 262 bp template. Purification was carried out using 'Qiaquick PCR purification kit' with 'MinElute' columns (Qiagen) following the manufacture protocol to ensure minimum contamination of the template with un-incorporated biotinylated primer. DNA concentration and purity was recorded prior to storing at -20°C (see 2.4.1.4).

Templates were coupled to Dynabead M-280 streptavidin coated magnetic beads (Invitrogen) following the manufacture protocol. Because a surplus is required relative to the template, initially 5 µL of homogenised stock bead solution (10 µg beads/µL in PBS pH 7.4, 0.1% BSA and 0.02% sodium azide) was aliquoted for binding an estimated 400 ng of template (binding 10 ng dsDNA per µg of beads). The beads are then washed in 1 mL of 1x binding/washing buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl), vortexed for 5 sec before collecting the beads next to a magnet for 1 min and discarding the supernatant. To

decrease potential RNase contamination, the beads were then washed twice with 1 mL of 'solution A' (DEPC treated 0.1 M NaOH, 0.05 M NaCl), then once with 1 mL of 'solution B' (DEPC treated 0.1 M NaCl), vortexing for 5 sec before collecting the beads next to a magnet for 1 min and discarding the supernatant after each wash. Finally the beads were re-suspended in 10 µL 2x binding/washing buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl) to achieve a final concentration of 5 µg beads/µL.

To couple the template, 10 µL of 200 nM biotinylated template was added to the 10 µL of beads, diluting the NaCl concentration from 2 M to 1 M for optimal binding. The template and beads were then incubated at room temperature with gentle rotation for 15 min. After incubation the coupled beads were washed 5 times in 1 mL of 1x binding/washing buffer, vortexing for 5 sec before collecting the beads next to a magnet for 2 min and discarding the supernatant after each wash remove any unbound template. Finally the beads were re-suspended in 10 µL DEPC treated H_2O to achieve a theoretical concentration of 200 nM coupled template before storing at -20°C.

2.9.2.2 Road-blocked *E. coli in vitro* transcription reaction

The road-blocked *in vitro* transcription reaction was carried out as above (see section 2.9.1) with some deviation (Chauvier *et al.* 2017). A total reaction volume of 200 µL was used, allowing 5 min elongation time after which the reaction was split into two aliquots of 100 µL each and placed next to a magnet for 1 min to collect the beads. The supernatant from one aliquot was collected and mixed in equal volume with 2x formamide stop buffer loading dye to capture released transcripts. To remove any un-bound transcripts, the beads were then washed 5 times with 100 µL of warmed 1x transcription buffer, mixing by gentle pipetting and collecting the beads next to the magnet for 1 min each wash. After washing, the beads were re-suspended in 100 µL of warmed 1x transcription buffer and mixed in equal volumes with 2x formamide stop buffer loading dye (see section 2.3). The second reaction aliquot served as a process control and was fully re-suspended before being mixed in equal volume with 2x formamide stop buffer loading dye. Reactions were stored at -20°C prior to use. Before loading for gel electrophoresis samples were denatured at 95°C for 10 min causing the template to uncouple from the beads and following denaturation samples were placed next to a magnet to separate any residual beads.

2.9.3 Sequencing urea polyacrylamide gel electrophoresis

Transcription reactions were separated using denaturing sequencing urea PAGE: 7 M urea, 1x TBE buffer and 10% (v/v final) of 40% 19:1 acrylamide:bis-acrylamide (Fisher Scientific) in 40 mL with dH_2O, cross-linked with 400 µL of 10% (w/v) APS and 40 µL TEMED. Gels were pre-run at 50 W constant, 50°C for 1 hour to warm the gel prior to loading. Once loaded, gels were run at 15 W constant using loading dye to estimate RNA migration. Once migrated gels were fixed with fixing solution of 20% acetic acid (v/v, Sigma), 20% methanol (v/v, Fischer Scientific) in dH_2O for 5 min before being vacuum dried on Whatman 3MM paper at 80°C for 1 hour. Once dried, gels were wrapped in Saran wrap to expose to a phosphor screen developed using a Typhoon FLA 9500 (GE).

2.10 *Mycobacterium tuberculosis* RNAP *in vitro* transcription system

Recombinant *Mycobacterium tuberculosis* RNAP holoenzyme was generously provided by Professor Sivaramesh Wigneshweraraj's group at the MRC Centre for Molecular Bacteriology and Infection, Imperial College London. Single-round *in vitro* transcription reactions were carried out as described above (see section 2.9.1). Halted TECs were assembled with the addition of 1x reaction buffer (20 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, 20 nM dsDNA template, 40 U 'RNasin plus' RNase inhibitor (Promega) and 400 nM *M. tuberculosis* holoenzyme 22.5 µL total with DEPC treated H_2O. Reactions were incubated 30°C for 5 min before adding 2.5 µL pre-warmed 10x halt mix (1500 µM GpU dinucleotide, 25 µM rGTP/rCTP with 7.5 µM rATP (Promega) and 0.4 µCi/µl 32P-αATP (3000 Ci/mmol 10 mCi/mL, Perkin Elmer). Reactions were then incubated at 37°C for 15 min. After incubation pre-warmed heparin (Sigma) was added to 200 ng/µL final to prevent additional transcription initiation in 45 µL with 1x buffer, incubating at 37°C for 5 min. Transcription elongation activated with the addition of 5 µL pre-warmed 10x elongation mix (2 mM of all rNTPs) at 200 µM final in 50 µL. Reactions extended at 37°C for 20 min before being stopped with 50 µL 2x formamide stop buffer loading dye. Time-course reactions were extracted as indicated. Reactions were stored at -20°C prior to use. Reactions were size separated by PAGE as described above (see section 2.9.3).

3 Re-annotation of *rpfB* translational regulation

3.1 Mis-annotation of mycobacterial ORFs

When considering regulatory nucleic acids, it is important to define the boundaries of the gene of interest on which the regulatory control will be exerted. In the case of *rpfB*, the riboswitch sequence lies upstream of the coding sequence within the 5'untranslated region. There is also evidence that supports the notion of an antisense sRNA that could regulate the gene (see section 1.10.4). For an accurate understanding of the impact of these regulatory features it is important to ensure the protein coding sequence (CDS) is correctly annotated. Much of the *Mycobacterium tuberculosis* genome is annotated using software; open reading frames (ORFs) may be annotated by identifying regions between start and stop codons, or using proximity of start codons to ribosome binding sites, which can often lead to mis-annotation of genes (Smollett *et al.* 2009).

Mycobacteria use alternative start codons more frequently than other bacterial species – as would be expected with a high G:C content (67%) genome – where it is estimated that in addition to ATG, coding sequences are initiated with GTG in 34% of genes and TTG in 5% (compared with 14% and 3% in *E. coli* respectively) (Blattner *et al.* 1997; Cole *et al.* 1998; DeJesus, Sacchettini & Ioerger 2013). Furthermore, many genes encoded by mycobacteria

are leaderless, with the transcription start site situated within 5 bp of the translation start site, and therefore initiate translation without an upstream canonical ribosome binding site (Cortes *et al.* 2013).

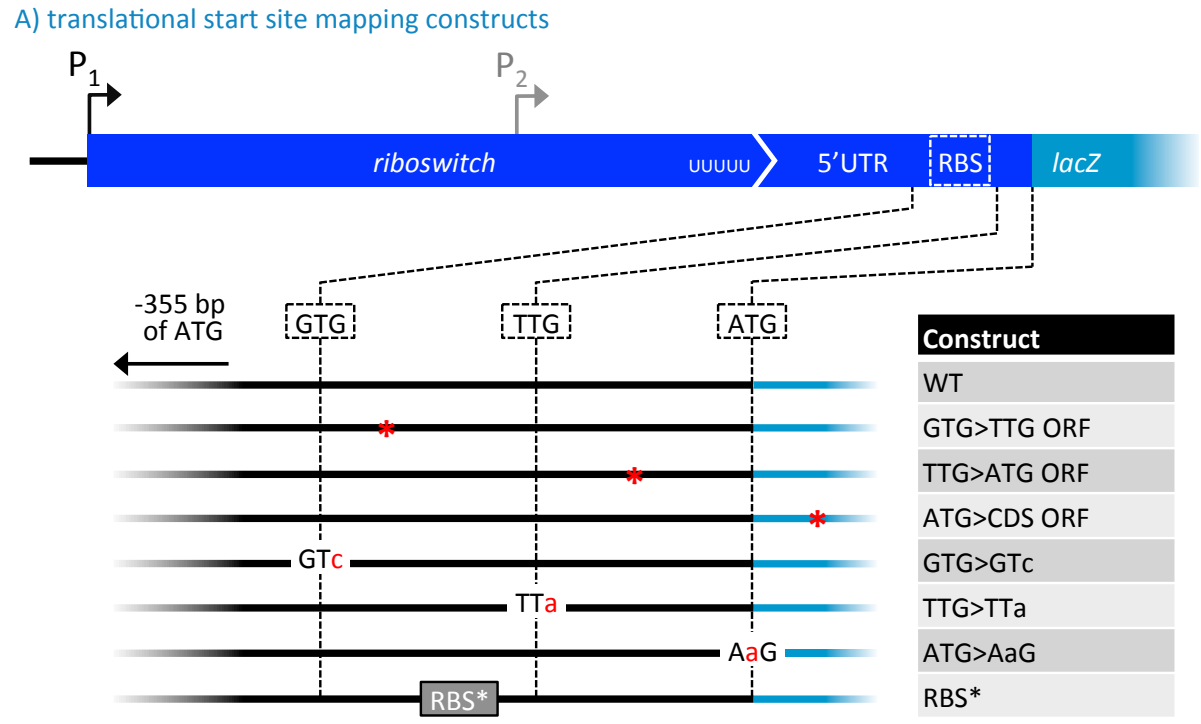
Correct ORF annotation can be achieved through translational start site mapping (TLSM), a method in which potential ORFs are fused with a reporter gene or tag protein and then individual loci are mutated to identify in which ORF the coding sequence starts (Smollett *et al.* 2009). In addition to the annotated ATG start codon, *rpfB* has a further two possible translational start sites upstream in *M. tuberculosis*. The most upstream is GTG followed by TTG and finally the annotated ATG start codon. Recently Sharma *et al.* (2015) reported the *rpfB* ATG annotated start codon as being the correct translation start site. Using a *lacZ* translational fusion of *rpfB* spanning -227 bp upstream of the ATG down to +122 bp, they demonstrated that by mutating ATG to AaG a significant loss of expression could be observed. However, the authors did not investigate the two additional potential start codons upstream of the annotated ATG start: TTG -36 bp upstream and GTG -72 bp upstream.

In the *Mycobacterium smegmatis rpfB* homologue (MSMEG_5439) the annotated start codon (residue 5522319) is further upstream than in *M. tuberculosis,* more proximal to the *M. tuberculosis* TTG rather than ATG (see Appendix Figure 10.I) (Kapopoulou, Lew & Cole 2011). If the start codon in *M. tuberculosis rpfB* were at TTG this would place the translational start site closer to the riboswitch stem-loop terminator as well as to the 5'end of the potential asRNA, which may serve to occlude a potential ribosome binding site. The work detailed in this chapter aimed to determine the translational start site of *rpfB* by TLSM and enabled correct construction of transcriptional and translational constructs to further elucidate regulation of the gene.

3.2 Frameshift reporters identify extended *rpfB* ORF

To investigate the potential start codons, a series of frameshift constructs were generated containing WT *M. tuberculosis* H37Rv *rpfB* 5'UTR spanning -355 bp upstream of the annotated ATG start codon down to +3 bp (including the ATG codon), this was then fused with *lacZ* (where the *lacZ* ATG had been removed). Frameshifts were introduced using separate deletions between the GTG and TTG, TTG and ATG codons and finally downstream of ATG within the *lacZ* CDS as a frameshift control. These constructs were transformed into *M. smegmatis* as a surrogate host, and expression was measured by β-galactosidase assay.

Protein extracts of *M. smegmatis* transformed with WT translational fusion showed a β-galactosidase activity of 37.59 ± 5.44 Miller units (MU)/mg total protein, representative of WT translational expression of *rpfB* (see Figure 3.I). When a frameshift mutation was introduced within the GTG ORF, between the GTG and TTG start codons, no significant change in expression relative to the WT was observed. However, when a frameshift was introduced within the TTG ORF, between the TTG and ATG start codons, a 12-fold significant decrease in expression to 3.01 ± 2.67 MU/mg total protein was observed. A similar result was also seen when a frameshift was introduced with the ATG ORF. Taken together these results indicated the *rpfB* ORF extends up to the TTG start codon.



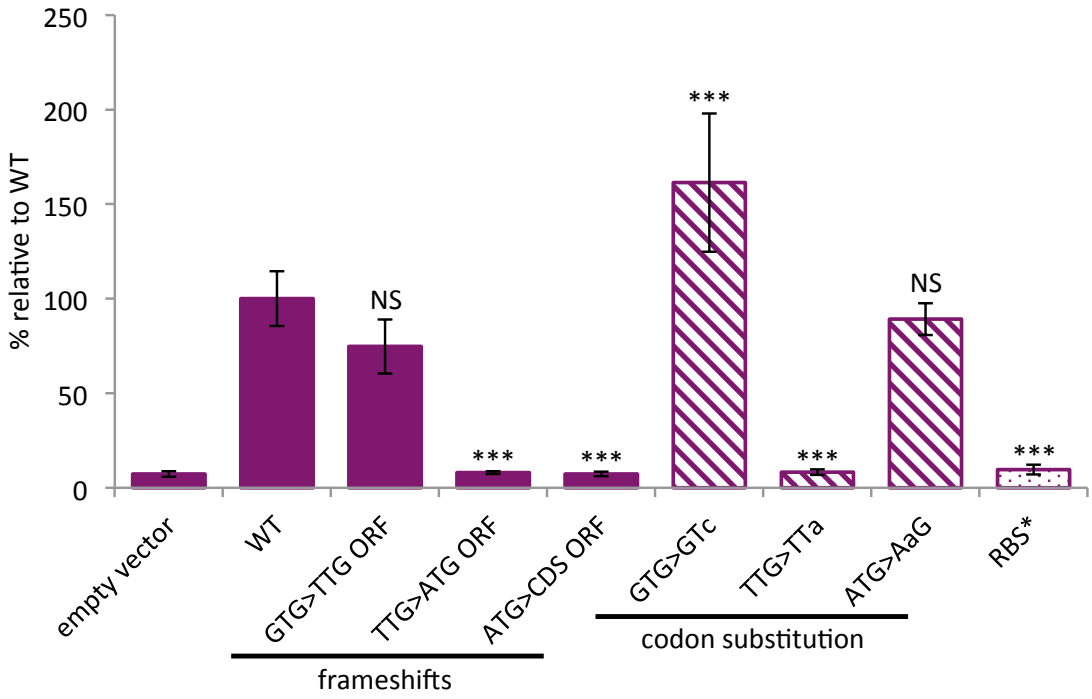


Figure 3.I translational start site mapping reporters
Diagram (A) depicts mutations introduced to each construct, where '*' indicates basepair deletions, red bases indicate substitutions and 'RBS*' indicates ribosome binding site mutation. Chart (B) shows results of β-galactosidase assay of protein extracts from transformed *M. smegmatis*. Mean of biological replicates (n=6) relative to WT. Statistical significance relative to WT calculated by one-way ANOVA with Tukey post hoc analysis in IBM SPSS ±1 standard deviation, NS $P > 0.05$, *** $P \leq 0.001$.

3.3 Start codon substitution reporters confirm correct *rpfB* start codon

The frameshift constructs demonstrated the *rpfB* ORF likely extended further upstream than the annotated ATG start codon, up to the TTG start codon. To confirm this finding, three more mutants were derived from the WT translational reporter. Separately each start codon was substituted with a synonymous non-start codon: GTG substituted with GTc and TTG substituted with TTa. Synonymous codon substitution is not possible for fMet so ATG was substituted with AaG to repeat the mutation used by Sharma *et al.* (2015).

When TTG was substituted with TTa a 12-fold significant reduction in expression to 3.11 ± 0.55 MU/mg total protein was observed relative to the WT fusion, very similar to that observed in the TTG ORF frameshift (see Figure 3.I). Mutation of the ATG start codon to AaG did not result in any significant change of expression. Mutation of the GTG start codon to GTc resulted in a significant increase in expression to 60.68 ± 2.67 MU/mg total protein, 1.6-fold greater than the WT; this was surprising although it may possibly be attributed to dis-regulation of other proteins binding (see section 3.5). Despite this, only through mutation of the TTG start codon was translational expression decreased to a level similar to that of the empty vector. Collectively the results corroborate those of the frameshift mutants (see section 3.2) and strongly indicate the translational start codon *M. tuberculosis rpfB* is TTG rather than ATG.

3.4 *rpfB* ribosome binding site validation

With the translational start codon of *rpfB* re-annotated to TTG it was necessary to investigate if *rpfB* required a ribosome binding site for translation. No obvious RBS was within canonical proximity of the TTG start codon, however this is not unusual in mycobacteria, which generally have a poor RBS consensus but a bias for ~6 base long purine rich regions starting at around -8 to -11 bp upstream of the start codon (DeJesus, Sacchettini & Ioerger 2013). An ambiguous purine rich sequence of 'GAGGTCGGGGA' was identified -20 to -10 bp upstream of the TTG start codon. This sequence was found to have some complementarity to the 16S rRNA 3'end, with a 5 base consecutive region capable of basepairing. To investigate if this sequence constituted a potential Shine-Dalgarno (SD) sequence it was mutated in the WT translational fusion described above, substituting all the

purines with pyrimidines (GAGGTCGGGGA to ctccTCcccct) to disrupt ribosome binding (Shine & Dalgarno 1975).

Upon mutation of the potential RBS, expression of β-galactosidase in transformed *M. smegmatis* decreased significantly relative to the WT (see Figure 3.I). A 10-fold decrease in expression to 3.61 ± 0.98 MU/mg total protein was observed. As the frameshift mutants described previously demonstrated this region to be outside of the CDS, this decrease in expression was attributed to reduced ribosome binding upstream of the translational start site within this region. This further supported the notion that TTG was the correct translational start site as this RBS is within a feasible proximity to the start codon (DeJesus, Sacchettini & Ioerger 2013).

3.5 Chapter discussion

Through translational start site mapping, the β-galactosidase assay results derived from the ORF frameshift mutants shown in Figure 3.I demonstrated that the *rpfB* CDS extends further upstream of the ATG start codon than previously annotated. The frameshift within the TTG to ATG ORF was the most upstream ORF to demonstrate reduced β-galactosidase expression relative to the WT. This indicated that the CDS of *rpfB* extended up to the TTG start codon. The exchange of the potential start codons with synonymous non-start codons corroborated this finding, as once again the TTa mutation was the most upstream mutation to affect expression. Although other start codons may lie upstream of the codons investigated in this study, none could allow *rpfB* translation; the next codon upstream of the GTG is a stop codon at -73 bp relative to the previously annotated ATG. These results indicate the TTG start codon of *M. tuberculosis rpfB* (Rv1009) is the site at which translation begins. This allows a greater conservation with the annotated start in the *M. smegmatis* homologue (MSMEG_5439). Therefore, it is strongly recommended the CDS of *M. tuberculosis rpfB* should be re-annotated and extended up to the TTG codon.

Recently Sharma *et al.* (2015) reported mutation of the ATG start codon of *M. tuberculosis rpfB* to AaG in a translational fusion resulted significant in loss of expression. This effect was not observed in the present study, in which the same substitution in the translational fusion resulted in no significant change in expression relative to the WT. The cause for this discrepancy is unknown; however it is clear that the constructs employed in either study

were not identical. The reporter used by Sharma *et al.* was fused with *lacZ* +122 bp downstream of the ATG start codon, whereas reporters in the present study were fused with *lacZ* at the ATG codon itself (+3 bp). It has previously been reported that the N-terminal of RpfB encodes a possible export signal peptide from residues 1-29, translated from +1 to +87 bp downstream relative to the ATG start codon (Ruggiero *et al.* 2009; Lew *et al.* 2011). The presence of an export signal peptide in fusion with β-galactosidase might have interfered with the protein extraction process in the Sharma *et al.* β-galactosidase assays. This signal sequence could not have affected results in the present study as it was not encoded by the reporters as none of the annotated *rpfB* CDS was included in the fusions. Furthermore, the constructs used in this study were further validated by observing reporter expression in transformed *M. tuberculosis* where they showed perfect agreement with the *M. smegmatis* β-galactosidase assay of Figure 3.I.

One intriguing finding was the discrepancy in expression between the frameshift in the GTG to TTG ORF compared with GTc synonymous non-start codon substitution. Introducing a frameshift within the GTG ORF resulted in an insignificant decrease in expression to 28.10 ± 5.38 MU/mg total protein relative to the WT (Figure 3.I). Conversely, substituting the GTG start codon with a synonymous non-start GTc codon resulted in a significant increase in expression to 60.68 ± 2.67 MU/mg total protein, 1.6-fold greater than the WT. The TLSM results indicate the GTG ORF to be outside of the *rpfB* CDS and none of the mutations to the GTG ORF were within the identified RBS. How could this area have affected expression of the reporters?

One hypothesis is that this area constitutes an Lsr2 binding site, a nucleoide-associated protein, and that these mutations coincidently may have altered the affinity with which Lsr2 binds this area to affect expression (Galagan *et al.* 2010; Gordon *et al.* 2010; Gordon 2013). Alternatively, it may be that mutations within this region interfered with the two-component response regulator, MtrA, which Sharma *et al.* found to regulate *rpfB*. However, none the mutations made to the GTG ORF in this study lie within the binding footprint of MtrA (-144 to -88 bp upstream of the ATG) that they describe. It may also be possible that the mutations made within the 5'UTR may have affected the stability of the mRNA transcript to cause changes in transcript secondary structure or susceptibility to RNase degradation, manifesting as variations in β-galactosidase expression (Thomas 2001).

Despite this, validation of these theories was without the scope of identifying the translational start site, and the discrepancy in the trends of the GTG codon mutations should not detract from the overall data, which indicate TTG as the correct translational start site.

3.6 Chapter conclusion

In conclusion, these results indicate that translational start site of *rpfB* has been incorrectly annotated. The CDS of *rpfB* should therefore be re-annotated, extending up to and including the TTG start codon. This finding is further supported by the identification of a region complementary in sequence to the 16S rRNA 3'end, which likely constitutes a ribosome binding site. As the re-annotation of the *M. tuberculosis rpfB* (Rv1009) shares extended conservation with the annotated start in the *M. smegmatis* homologue (MSMEG_5439) it is possible that other mycobacterial *rpfB* homologues with strong sequence identity may also be mis-annotated.

4 *rpfB* has a regulatory relationship with *ksgA*

4.1 KsgA is a universally conserved methyltransferase

Ribosome biogenesis is a crucial process that has evolved within all domains of life. Biogenesis of ribosomal subunits is complex, involving the contribution of multiple proteins, ribosomal RNAs (for prokaryotic ribosomes the 23S and 5S in the large subunit, 16S in the small subunit) and modifications that have evolved divergently across the domains of life (Mangat & Brown 2008). Despite this, there are some ribosomal modifications that are universally conserved, such as methylation of two adjacent adenosine residues in the 3'terminal helix (helix 45) of the 16S ribosomal RNA (rRNA) (Xu *et al.* 2008; Demirci *et al.* 2010). This modification is important for stable folding of the 16S rRNA 3'terminus which constitutes the Shine-Dalgarno (SD) sequence (Shine & Dalgarno 1975; Kaberdina *et al.* 2009).

In 1965 a *Streptomyces* strain (*S. kasugiensis*) was isolated from the Kasuga shrine in Nara City, Japan (Umezawa *et al.* 1965). The isolate produced a novel aminoglycoside capable of inhibiting translation initiation in many bacteria. Subsequently, antibiotic resistance was identified in *Escherichia coli* that lacked the previously described 16S rRNA methylation. The loss of methylation was attributed to mutations in a methyltransferase later named *ksgA*

(Lange *et al.* 2017). All domains of life encode a *ksgA* methyltransferase that is responsible for this methylation of adjacent adenosine residues in rRNA.

Whilst mutation of *ksgA* renders bacteria resistant to kasugamycin, deletion is not lethal but impacts ribosome biogenesis, translation efficiency and potential virulence as a fitness cost (Kaberdina *et al.* 2009; Kyuma *et al.* 2015). KsgA requires S-adenosyl-L-methionine (SAM) as a methyl group donor to methylate rRNA. Connolly, Rife & Culver (2008) found that a catalytically inactive mutant of KsgA, incapable of binding SAM, was unable to release un-methylated 16S rRNA. Results indicated not only an accumulation of immature small ribosomal subunits, lacking the rRNA methylations, but also that KsgA incapable of binding SAM remains bound to the small ribosomal subunit where the large subunit would interact (Xu *et al.* 2008). This indicates KsgA serves as a checkpoint for ribosome biogenesis, preventing immature un-methylated ribosomal subunits from forming the complete 70S ribosome (Connolly, Rife & Culver 2008; Mangat & Brown 2008). The role *ksgA* plays in ribosome biogenesis, and by extension protein synthesis, makes *ksgA* methyltransferases promising as potential therapeutic drug targets. Interfering with KsgA activity could be a very efficient method of inhibiting protein synthesis as part of a combined therapy.

In *Mycobacterium tuberculosis* there is a 28 bp overlap between the 3'end of the *rpfB* ORF (Rv1009) and the 5'end of the *ksgA* ORF (Rv1010) which is annotated by homology to the *E. coli ksgA* (Tufariello, Jacobs & Chan 2004). This is not an unusual grouping (see Figure 4.I), as many firmicutes and actinobacteria share a conserved arrangement of a *tatD* homologous deoxyribonuclease upstream of a cell wall hydrolase and a downstream *ksgA* methyltransferase (Ravagnani, Finan & Young 2005).

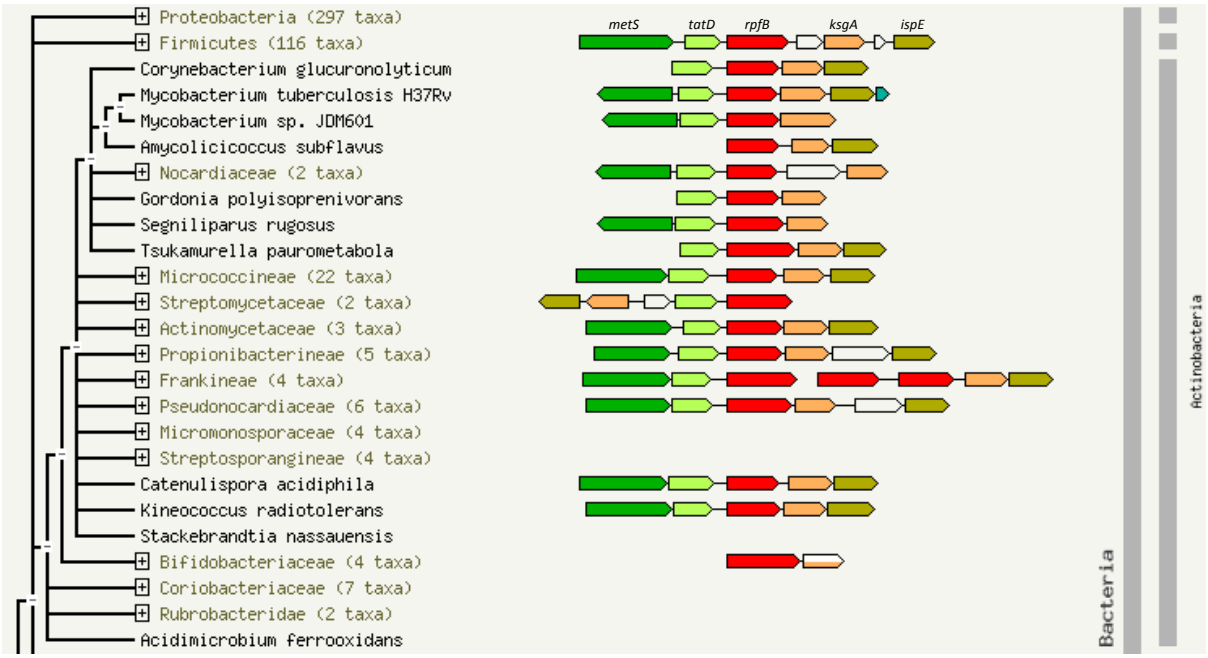


Figure 4.I *rpfB* gene neighbourhood
Gene neighbourhood diagram surrounding *rpfB* in actinobacteria generated in STRING (Szklarczyk *et al.* 2017). Starting upstream for *Mycobacterium tuberculosis*; *metS* methionine tRNA ligase in dark green, *tatD* deoxyribonuclease in light green, *rpfB* resuscitation promoting factor in red, *ksgA* methyltransferase in orange.

A conserved proximity of genes often indicates a regulatory or functional link referred to as "guilt-by-association", where it is beneficial to have genes grouped by function (Szklarczyk *et al.* 2017). In the case of *rpfB* and *ksgA* there is extensive conservation in proximity, which indicates a potential link between cell wall synthesis and protein synthesis (Ravagnani, Finan & Young 2005). The work detailed in this chapter aimed to determine the relevance of this link in regards to the regulation exerted by the *rpfB* riboswitch.

4.2 *rpfB* and *ksgA* form a bi-cistron

Returning to the transcription start site mapping published by Cortes *et al.* (2013) (see section 1.8) it was apparent that *ksgA* did not have any strong transcription start site (TSS) signal sequence of its own (see Figure 4.II). The only clear TSSs were those of *rpfB* P_1 and P_2 promoters, situated upstream. A hypothesis was developed that *rpfB* and *ksgA* are transcribed as an operon from the *rpfB* promoters.

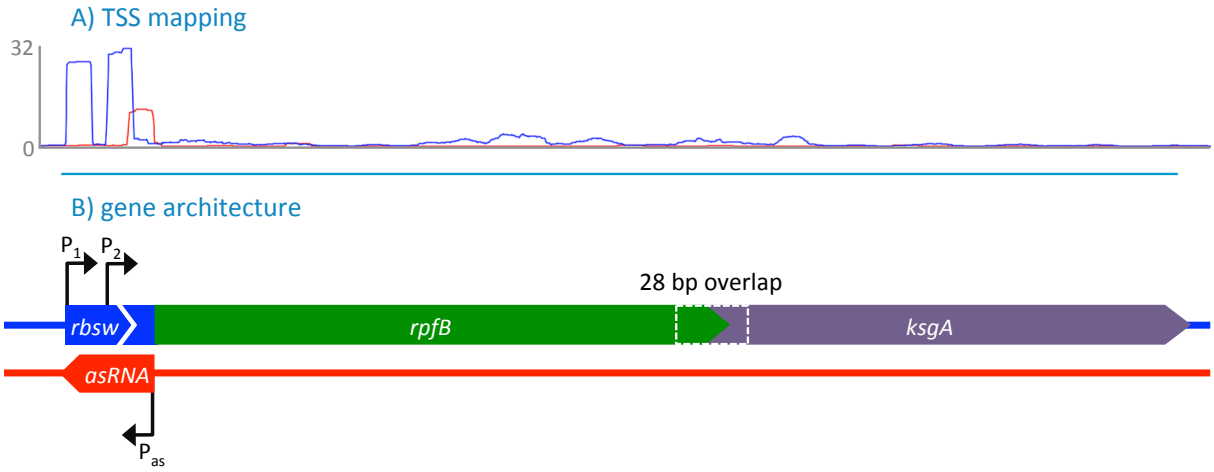


Figure 4.II transcription start site mapping of *rpfB* locus
Chart (A) shows global transcriptional start site analysis of sense (blue) and antisense (red) primary transcript abundance normalised to total reads in *M. tuberculosis* (Cortes *et al.* 2013). The bottom diagram (B) depicts the organisation of genetic elements in alignment with the chart above it.

cDNA was generated through reverse transcription of exponential phase *M. tuberculosis*, *Mycobacterium bovis* BCG and *Mycobacterium smegmatis* total RNA extracts. Genomic DNA (gDNA) was extracted as a control for *M. tuberculosis* and *M. bovis* BCG, which share 100% sequence identity for the *rpfB-ksgA* locus. A separate extract was made for *M. smegmatis* in which *rpfB* and *ksgA* are proximal but do not have an annotated ORF overlap. cDNA and gDNA were subsequently PCR amplified, priming either side of the region between *rpfB* and *ksgA* (see section 2.7.5).

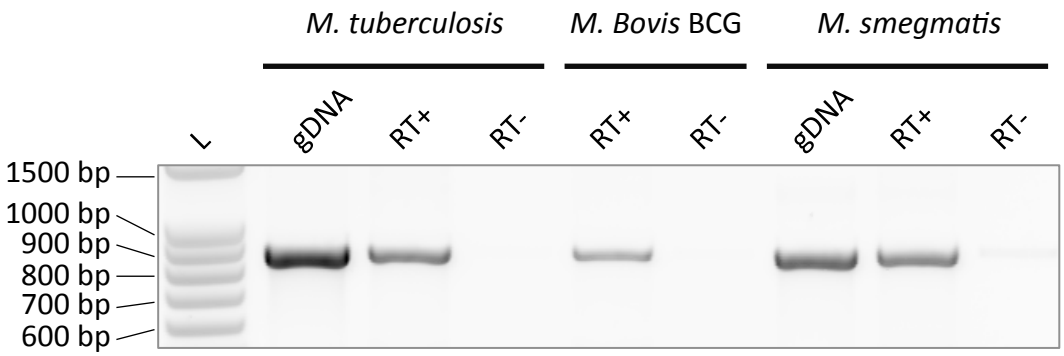


Figure 4.III RT-PCR analysis of *rpfB-ksgA*
RT-PCR analysis or the *rpfB-ksgA* overlapping region. From left to right: *M. tuberculosis* H37Rv gDNA followed by exponential phase cDNA, *M. bovis* BCG exponential phase cDNA, and *M. smegmatis* gDNA and exponential phase cDNA. Results show a single amplicon for *rpfB-ksgA* of ~850 bp, ± reverse transcriptase (RT). gDNA served as a positive control. Reactions primed with oligonucleotides 5.51 and 8.04 (see Appendix section 10.3).

Figure 4.III shows a single ~850 bp amplicon generated by reverse transcription PCR (RT-PCR) in all three species. This is strong evidence that *rpfB* and *ksgA* are co-transcribed as a bi-cistron from the *rpfB* promoters upstream. Such a transcriptional link between the two genes re-enforces the previously proposed functional link between cell wall synthesis and protein synthesis (Ravagnani, Finan & Young 2005). Moreover, for the pathogenic species such as *M. tuberculosis* and *M. bovis* BCG that encode the *rpfB* riboswitch upstream of the bi-cistron, this transcriptional link is further re-enforced by the regulation exerted by the riboswitch.

4.3 *rpfB* and *ksgA* are not co-translated

ksgA has no apparent proximal purine rich ribosome binding site of its own. As described previously (see section 3.3) this is not unusual for mycobacteria, which generally have a poor RBS consensus (DeJesus, Sacchettini & Ioerger 2013). However, following the discovery of the riboswitch regulated co-transcriptional link described above, it was hypothesised the regulatory link between the proteins may extend to the translational level.

To test this theory, three constructs were made in fusion with the heterologous constitutive PCLI promoter (Gonzalez-y-Merchand *et al.* 1997); spanning from -1120 bp (inclusive of native *rpfB* RBS), -75 and -25 bp upstream of the *ksgA* start codon down to +150 bp in fusion with *lacZ* (see Figure 4.IV). For all three constructs, transformed *M. smegmatis* extracts exhibited reporter expression when assayed for β-galactosidase activity. The -1220 bp fusion included the native *rpfB* RBS upstream of the fusion as described previously (see section 3.3). However, for the -75 and -25 bp fusions, translation was indicated by the expression of β-galactosidase, despite the absence of any RBS sequence. This demonstrated translation of *ksgA* is possible in the absence of the *rpfB* RBS, and therefore that the two proteins are not co-translated.

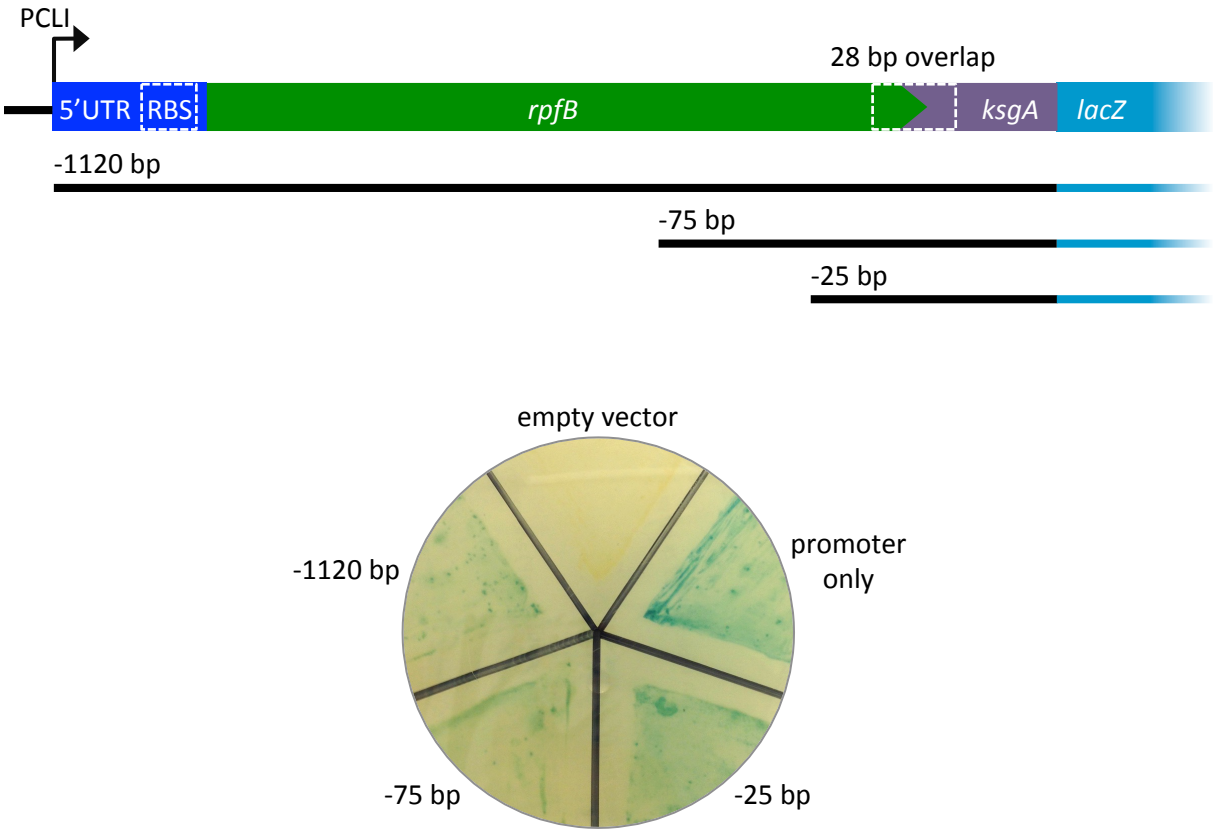


Figure 4.IV *ksgA* translational reporters
Translational fusions of *ksgA* upstream regions. Top diagram depicts the fragments of the *rpfB-ksgA* bi-cistron fused between the heterologous constitutive PCLI promoter TSS upstream and downstream *lacZ* reporter. The image beneath shows the expression of β-galactosidase on X-gal LB agar in transformed *M. smegmatis*, with controls of empty vector and promoter only with no insert and native *lacZ* RBS.

4.4 Chapter discussion

The highly conserved proximity of *rpfB* and *ksgA* throughout actinobacteria and firmicutes has been identified previously with theories of a regulatory or functional link between cell wall synthesis and protein synthesis (Ravagnani, Finan & Young 2005). In this study, through RT-PCR, such a theorised regulatory link has been further substantiated. The amplicons shown in Figure 4.III indicated that *rpfB* and *ksgA* form a single transcript as a bi-cistron in *M. tuberculosis*, *M. bovis* BCG and *M. smegmatis*. Such an operon may also be present in other actinobacteria where the proximity of the genes is conserved. Moreover, in the pathogenic mycobacteria that encode the *rpfB* riboswitch, this co-transcriptional link has an additional layer of regulation exerted by the riboswitch upstream.

The biological relevance of this link remains to be understood. For mycobacteria particularly, there is a clear inference of a regulatory link between resuscitation from dormancy by RpfB cell wall hydrolase activity and the production of mature ribosomes by KsgA methylation of 16S rRNA. Potentially the two proteins may interact with each other; however investigation of any potential interaction was without the scope of this study regarding characterisation of the *rpfB* riboswitch. In future it could be worthwhile investigating if there is any interaction between RpfB and KsgA at the protein level. Although this would be would be unlikely as ribosomal maturation occurs within the cytoplasm and RpfB may be both cell wall anchored or secreted (Kana & Mizrahi 2010). Potentially there may be a secondary function that requires protein-protein interaction. Nevertheless, there is still potential for *ksgA* to have relevance to the riboswitch.

Structural models of the riboswitch predict that it has a default conformation of an intrinsic terminator (see section 1.9), making it a rarer example of a transcriptional 'on-switch' where it is predicted that ligand interaction would enhance transcriptional read-through (Zuker 2003; Tucker & Breaker 2005). Most riboswitches discovered to date employ a feedback loop, recognising a ligand generated by the gene or operon which they regulate (Dambach & Winkler 2009). For the *rpfB* riboswitch this would require a feedback produced by the cell wall hydrolase activity of RpfB. This is not unprecedented; in the case of the *ydaO* riboswitch, the secondary messenger c-di-AMP is sensed by the riboswitch as an indication of osmotic stress, but as an 'off-switch' (Block, Hammond & Breaker 2010; Nelson *et al.* 2013). In actinobacteria this c-di-AMP sensing riboswitch is commonly associated with cell wall enzyme regulation, and potentially *rpfA* in *M. tuberculosis*, therefore c-di-AMP may serve to modulate *rpfA* expression as an intermediate signal that is indirectly affected by the cell wall hydrolase activity (Nelson *et al.* 2013; Commichau *et al.* 2015).

Considering the *rpfB* riboswitch is a predicted 'on-switch', increased concentrations of ligand would theoretically increase the expression of the cell wall hydrolase RpfB and cause increased digestion of the cell wall peptidoglycan, a process for which tight regulation is crucial to cell survival. However, as the riboswitch has been demonstrated to regulate the *rpfB-ksgA* bi-cistron, regulatory control of *ksgA* expression must also be considered. As a ribosome maturation checkpoint, KsgA binds 16S rRNA and methylates two adjacent adenosine residues using SAM as a methyl donor to produce a mature 30S small ribosomal

subunit (Connolly, Rife & Culver 2008). In the absence of SAM KsgA does not release un-methylated 16S rRNA but remains bound to the 30S small ribosomal subunit at the interaction plane of the 50S large ribosomal subunit, precluding the assembly of immature 70S ribosomes.

SAM is used as a major methyl group donor in bacteria for a range of cellular processes; 16S rRNA methylation, biosynthesis of cell envelope mycolic acids and adhesion modifications (Parveen & Cornell 2011). Mycobacteria have been found to tightly control SAM synthesis, encoding multiple SAM synthases and methyltransferases (Cole *et al.* 1998; Berger & Knodel 2003). Increased concentrations of SAM could certainly be used as a signal indicative of suitable growth conditions for both resuscitation from dormancy and protein synthesis (Batey 2011). Indeed, SAM is one of the most common riboswitch ligands, with an extended superfamily of riboswitches capable of detecting either SAM or its un-methylated derivative S-adenosyl-L-homocysteine (SAH) (Price, Grigg & Ke 2014). In addition to the fact that in the absence of SAM, KsgA effectively inhibits the formation of functional ribosomes, it is feasible either SAM or its precursors/derivatives may serve as a ligand for the riboswitch, a hypothesis that is further investigated in this study (see section 7.2).

Whilst a strong regulatory link was inferred by the discovery of the *rpfB-ksgA* bi-cistron, the translational fusions described in Figure 4.IV demonstrated there was no translational link between the two genes. Transformed *M. smegmatis* were capable of translating the *ksgA-lacZ* reporter fusion in the absence of the native *rpfB* ribosome binding site upstream, despite any clear RBS within proximity of the *ksgA* start codon. Mycobacteria do not have a strong RBS consensus (DeJesus, Sacchettini & Ioerger 2013), however the upstream region of *ksgA* is relatively purine rich, therefore it may still be possible for the ribosome to bind in this region to translate *ksgA*.

Recently Yamamoto *et al.* (2016) described a novel method of translation initiation where assembled 70S ribosomes may initiate translation of ORFs without any proximal RBS by means of “70S-scanning”. Following translation termination, the 70S ribosome does not necessarily dissociate, but may freely translocate along the mRNA away from the stop codon, scanning for another start codon to initiate translation again. Evidence shows this method of translation initiation is widely conserved in all domains of life and has been

shown to be particularly effective within bi-cistronic mRNAs. Potentially this 70S-scanning mechanism could allow translation of *ksgA* in the absence of a RBS (Yamamoto *et al.* 2016).

Finally, the discovery of a riboswitch regulated *rpfB-ksgA* bi-cistron not only links the two genes, but increases the impact of the upstream riboswitch. With the developing healthcare threat of multi-drug resistant tuberculosis and a limited availability of effective antimycobacterial drugs, there is a growing need for novel therapeutic targets in *M. tuberculosis* (Hett & Rubin 2008). Both resuscitation from dormancy and ribosomal assembly processes have previously been identified as potential targets (Mangat & Brown 2008; Seidi & Jahanban-Esfahlan 2013). Not only has this work has identified a riboswitch that regulates the transcription of both *rpfB* and *ksgA* – proteins implicit in to both these processes – but the riboswitch itself is highly conserved in pathogenic mycobacteria. Riboswitches themselves make excellent therapeutic targets for rational drug design; they are well conserved between species and have evolved over thousands of years to identify a single cognate ligand with a high specificity and affinity (Blount & Breaker 2006). Although the current understanding of the biological process of resuscitation is limited, there can be no doubt that it contributes towards the development of an active infection from dormant bacteria. The *rpfB* riboswitch – capable of regulating two potential therapeutic targets related to resuscitation and ribosome maturation – could offer a highly specific target for rational drug design of antimycobacterials for a disease which is very much in need of novel therapeutics (Kana *et al.* 2014).

4.5 Chapter conclusion

In conclusion, these results further substantiate a functional link between resuscitation from dormancy and ribosome maturation. The discovery of a conserved mycobacterial *rpfB-ksgA* bi-cistron suggests that comparable operons may exist in other actinobacteria and firmicutes that encode genes with homologous function and proximity. Moreover, this co-transcriptional link offers potential insight into candidate ligands associated with *rpfB* or *ksgA* which could be recognised by the transcriptional riboswitch upstream and exploited for rational drug design of novel therapeutics.

5 Riboswitch-mediated transcription attenuation *in vivo*

5.1 RNA-seq reveals potential riboswitch upstream of *rpfB*

Global transcription start site mapping (TSSM) of *Mycobacterium tuberculosis* indicates that there are large 5'UTR upstream regions to some *rpf* mRNAs (Cortes *et al.* 2013). By mapping primary 5'triphosphorylated transcripts to the *rpfB* locus, two transcription start sites (TSSs) were identified. These TSSs were used to annotate two putative promoters (see Figure 5.I) designated P_1 and P_2, which have sequence identity similar to the SigA promoter -10 element 'TANNNT' consensus (Cortes *et al.* 2013). Transcripts initiated at the P_1 TSS may be expected to yield a large 5'UTR characteristic of a *cis*-acting regulatory RNA, encoding a potential intrinsic terminator, a stem-loop immediately followed by a poly(U) tract downstream (see section 1.9). Intrinsic termination upstream of the *rpfB* coding sequence (CDS) would prevent expression of RpfB, therefore it was hypothesised the proposed intrinsic terminator might constitute an expression platform of a riboswitch.

Riboswitch-mediated regulation of *rpfs* is not unprecedented as a *ydaO* riboswitch homologue has already been described within the *rpfA* 5'UTR (Block, Hammond & Breaker 2010; Arnvig & Young 2012; Nelson *et al.* 2013; St-Onge *et al.* 2015). RNA-seq data shown in

Figure 5.I indicates there may be processing of the 5'UTR beyond the potential intrinsic terminator, with transcript abundance increasing downstream of the poly(U) tract.

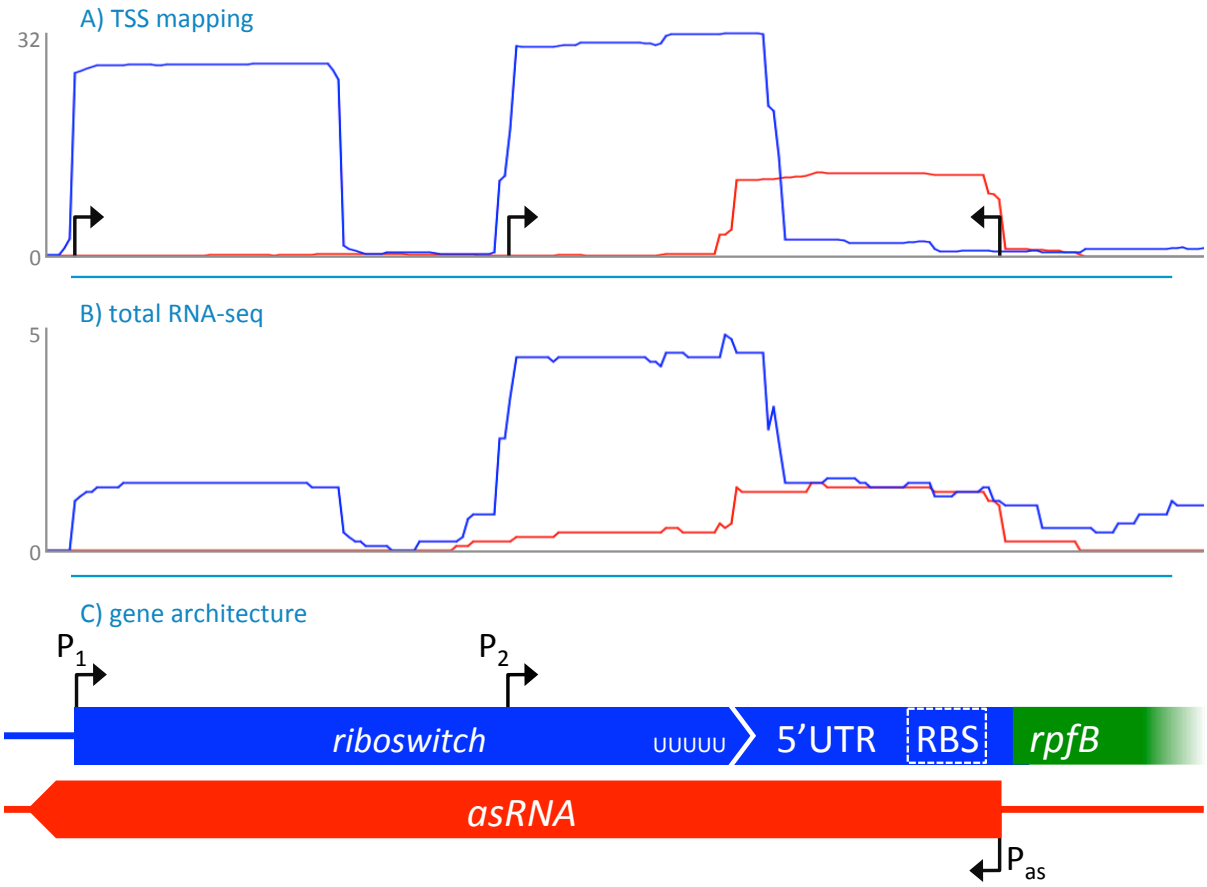


Figure 5.I diagram of *rpfB* RNA expression
Chart (A) shows global transcriptional start site analysis of sense (blue) and antisense (red) primary transcript abundance in *M. tuberculosis* (Cortes *et al.* 2013). Chart (B) shows the RNA-seq transcript abundance of sense (blue) and antisense (red) transcripts normalised to total reads (Cortes *et al.* 2013). Diagram (C) shows the organisation of genetic elements in alignment with the charts above it. Black arrows indicate putative transcription start sites.

The *M. tuberculosis* RNA-seq data Cortes *et. al* (2015) analysed was a representation of the entire transcriptome. To further understand individual components that contribute to the regulation of the *rpfB* locus would therefore require a more targeted approach. The work detailed in this chapter aimed to characterise the regulatory function of the riboswitch *in vivo*.

5.2 *rpfB* transcription attenuation is abundant in exponential growth

In the present study, it was hypothesised that it might be possible to detect riboswitch-mediated attenuation through northern blotting with a riboswitch complementary ribo-probe. Accordingly, total RNA was extracted from *M. tuberculosis* during both exponential and stationary growth phases, size separated by polyacrylamide gel electrophoresis (PAGE) and blotted onto a nylon membrane to be probed (see section 2.8).

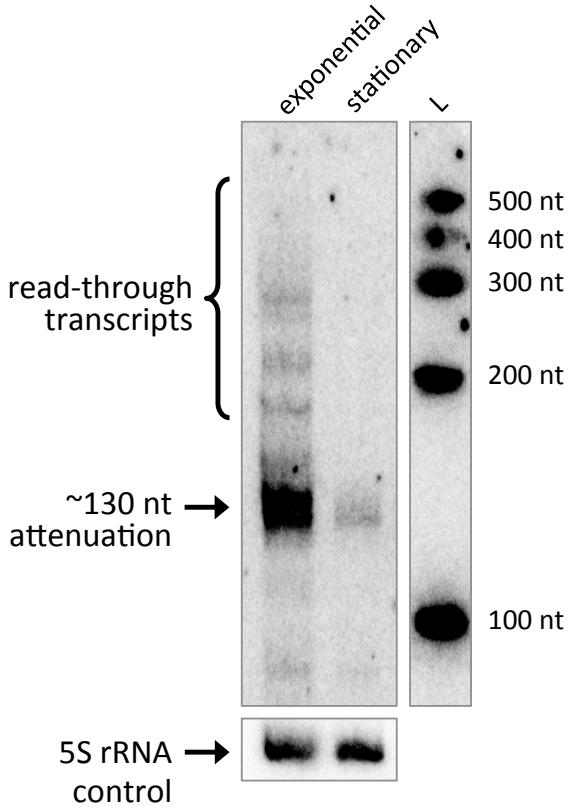


Figure 5.II northern blot of exponential and stationary growth phases
Northern blot of total RNA was extracted from *M. tuberculosis* H37Rv during both exponential and stationary growth phases (see section 2.2.2). RNA was probed within the *rpfB* P_1 5'UTR (oligonucleotide 5.22) and for 5S rRNA as a loading control (oligonucleotide 1.48, see Appendix section 10.3).

An absence of P_1 transcription in stationary growth phase was demonstrated by northern blot in Figure 5.II. The length of the *rpfB* riboswitch transcript from the P_1 TSS to the end of the intrinsic terminator poly(U) tract is predicted to be approximately 130 nt, depending on the position at which transcription attenuates. Therefore the ~130 nt transcript detected using the riboswitch complementary ribo-probe likely represented riboswitch attenuated transcript, with the smear of larger transcript indicating transcriptional read-though.

5.3 Promoter element reporter fusions

As the P_2 promoter is situated within the riboswitch sequence (see Figure 5.I), investigation of the P_1 and P_2 promoters separately would require truncation of the riboswitch. Therefore, a reporter system was devised to explore promoter function within the context of the riboswitch. A translational reporter fusion was made (identical to that employed for mapping of the translation start site (see section 3.2), containing WT *M. tuberculosis* H37Rv *rpfB* 5'UTR spanning -316 bp upstream of the re-annotated TTG start codon, with the first 14 codons of *rpfB* CDS fused with *lacZ* (see Figure 5.III, A).

The proposed -10 elements were mutated from 'TANNNT' to 'cANNNc' (see section 2.4.7), to make them less amenable to promoter melting. Mutations were made separately to the P_1 and P_2 -10 elements, changing 'TAGGGT' to 'cAGGGc', and 'TACCGT' to 'cACCGc' respectively. A third construct was made where both P_1 and P_2 were disrupted to scrutinise if any other regions capable of promoting transcription were encoded within the *rpfB* locus.

In addition to investigating the sense strand promoters, another reporter was constructed to explore potential P_{as} asRNA promoter activity. A transcriptional fusion was made spanning -100 bp upstream of the P_{as} TSS down to +2 bp; this was fused to a *lacZ* reporter in the absence of a vector encoded promoter (see Figure 5.III, B). Constructs were transformed into *M. tuberculosis* and plated onto X-gal agar to assess promoter activity qualitatively as represented by β-galactosidase expression.

The WT reporter shown in Figure 5.III corroborated previously observed expression in transformed *M. smegmatis* (see section 3.2). Compared to the empty vector, the WT reporter showed substantial expression. Where the P_1 promoter had been compromised there was diminished expression of the β-galactosidase reporter. Surprisingly, the P_2 promoter substitution resulted in an increase in expression, perceived to be exhibiting more β-galactosidase activity than in the WT. Mutation of both promoter elements resulted in a distinct lack of expression equal to the empty vector indicating no other regions capable of promoting transcription within the fused *rpfB* 5'UTR fragment.

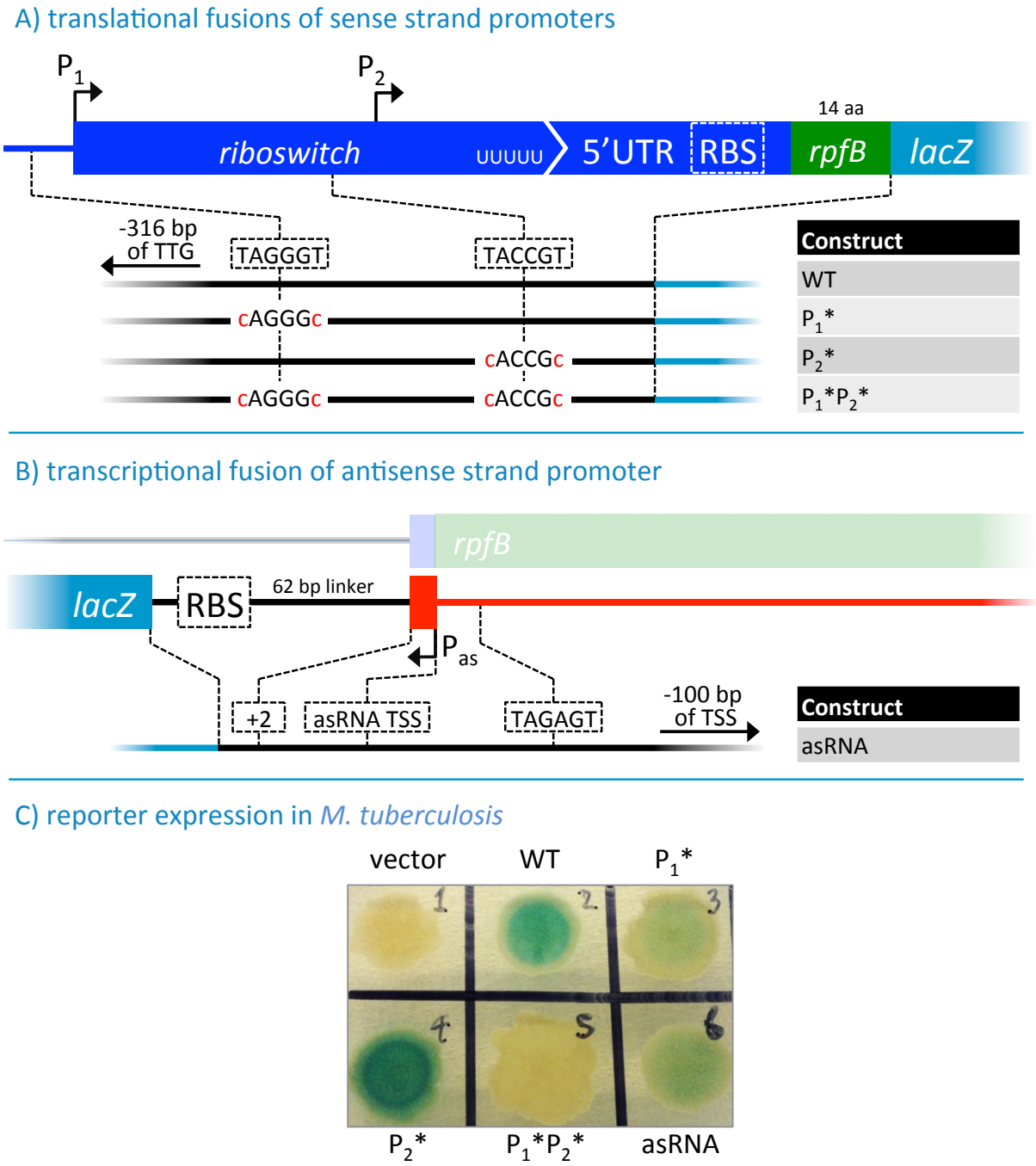


Figure 5.III promoter reporters

β-galactosidase reporter fusions of *rpfB* promoters. Diagram (A) depicts the construction of the translational fusion employed to investigate the sense strand promoters P_1 and P_2, '14 aa' denotes the first 14 codons of *rpfB* included in the fusion. Diagram (B) depicts the construction of the transcriptional fusion used to investigate the antisense strand promoter P_{as}, drawn in a 3' to 5' direction in reference to the sense stand shown faded above it. Image (C) shows the expression of the β-galactosidase reporter in transformed *M. tuberculosis* where blue colour indicated promoter activity. From greatest to least, perceived activity was recorded as $P_2^* \geq$ WT > asRNA > P_1^* > $P_1^*P_2^*$ = empty vector.

Regarding the antisense reporter, a weak level of expression was observed, further substantiating the Cortes *et al.* (2013) TSSM where a weak antisense signal was also indicated. Collectively results indicated the P_2 promoter makes an unsubstantial contribution to *rpfB* expression during exponential growth (as demonstrated by northern blot, Figure 5.II) or within a mixed population on agar (Figure 5.III). Consequently the majority of expression appears to be driven by the P_1 promoter and subject to riboswitch regulatory control.

5.4 Mapping of riboswitch attenuated transcripts

The 3'termini of transcripts were mapped to identify the location of potential riboswitch-mediated attenuation. Rapid amplification of cDNA ends (RACE) was adapted from Gehart, Wagner and Vogel (2008). Briefly, *M. tuberculosis* exponential phase total RNA was tailed by 3'polyadenylation to identify the 3'teminal, reverse transcribed, PCR amplified then cloned and sequenced (see section 2.7.3).

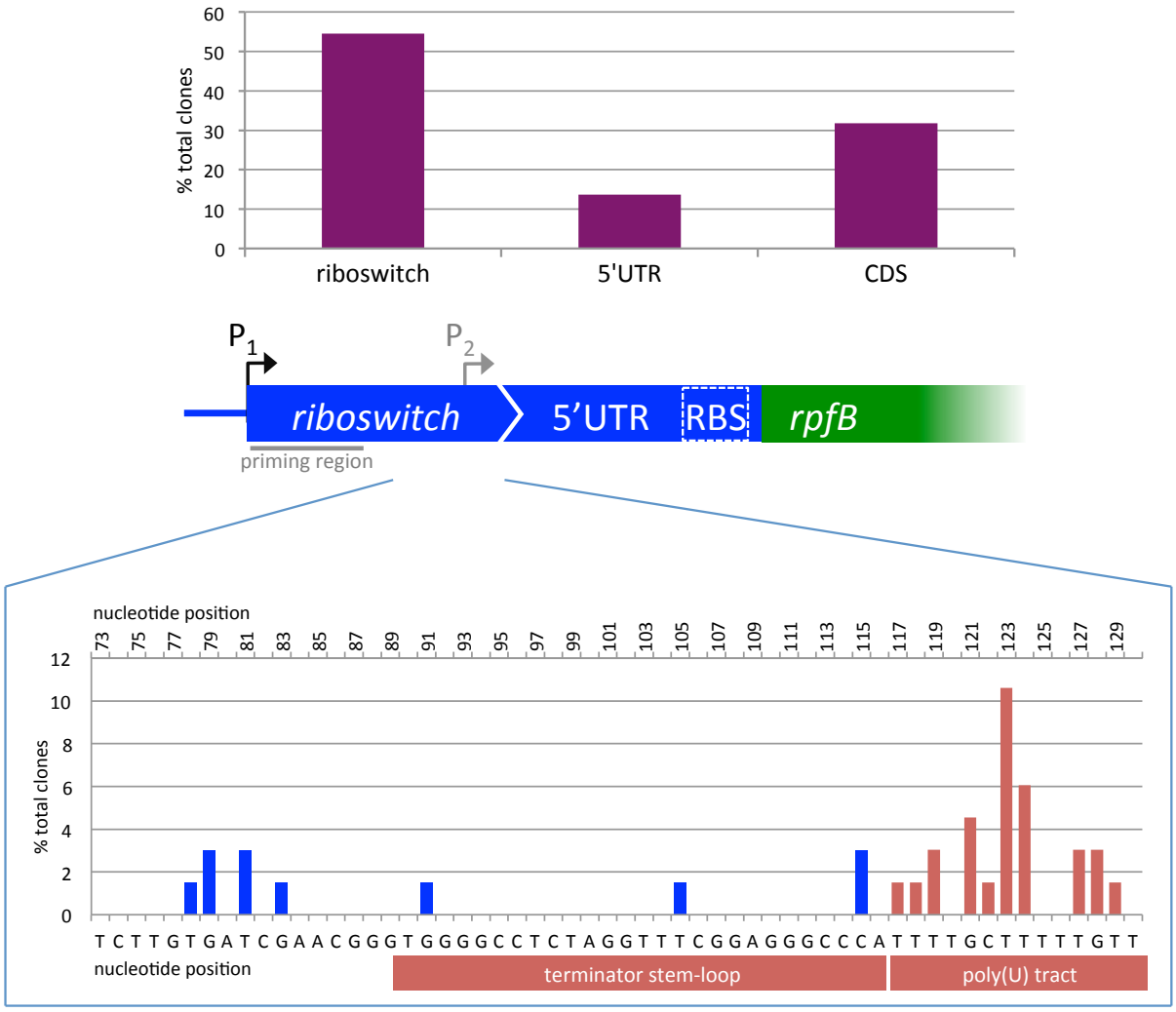


Figure 5.IV 3'RACE mapping
Diagram shows mapped 3'termini of *rpfB* transcripts from *M. tuberculosis* H37Rv exponential growth phase total RNA. Top chart shows total 3'termini mapped within the *rpfB* locus. Bottom chart shows 3'termini mapped within the riboswitch at single nucleotide resolution relative to the P_1 TSS.

A total of 66 unique clones were sequenced, of these 54.54% (36/66) mapped to the riboswitch region, with 42.42% (28/66) mapping specifically to the intrinsic terminator poly(U) (see Figure 5.IV). Within the poly(U) tract itself 10.61% (7/66) mapped to U123 (relative to the P_1 TSS), with multiple other clones mapping elsewhere within the poly(U) tract (see Figure 5.V). The most downstream poly(U) tract attenuation event was mapped to U129 (1.52%, 1/66), which was therefore used as evidence to annotate the riboswitch poly(U) tract as far as U130, the final uracil encoded within a relatively un-interrupted stretch. Downstream of the riboswitch, 31.81% (21/66) mapped to *rpfB* CDS, with a relatively high 7.58% (5/66) mapping to G179, the last nucleotide of the re-annotated TTG start codon (see section 3.2), indicating potential start codon proximal transcriptional pausing (see Appendix Figure 10.IV).

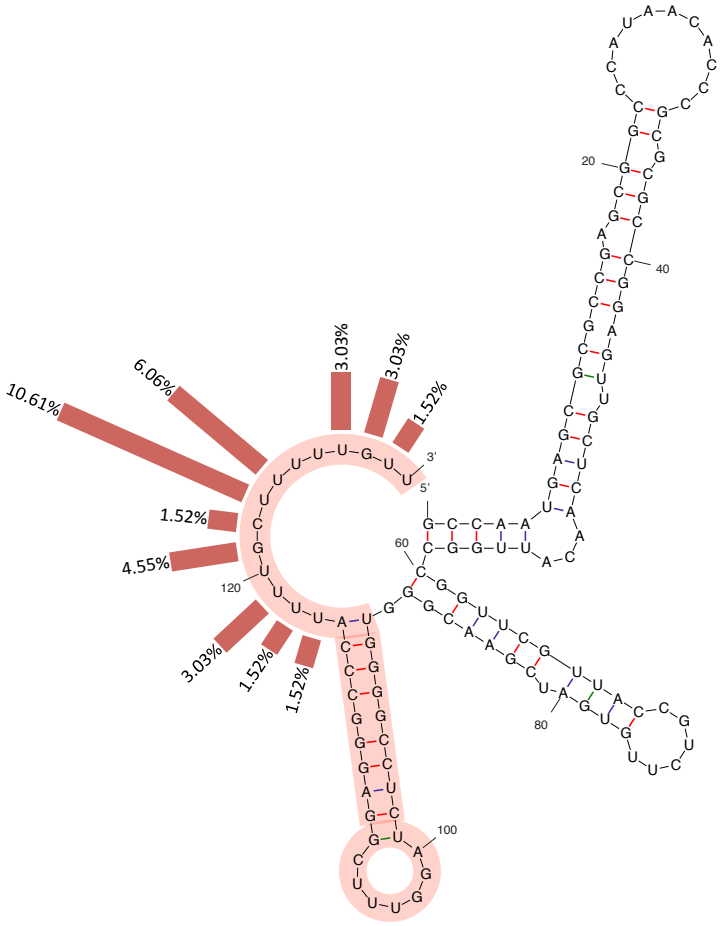


Figure 5.V mapped riboswitch attenuation
A graphical representation of the 3'RACE results mapped onto the riboswitch structure as predicted by Mfold (Zuker 2003). Red bars represent the percentage of total mapped clone sequences at each nucleotide of the riboswitch poly(U) tract.

The results of the northern blot (Figure 5.II) and 3'RACE (Figure 5.IV) both demonstrated riboswitch-mediated attenuation at the potential intrinsic terminator. Multiple attenuation events mapped to the poly(U) tract itself, indicative of intrinsic termination at these sites. In both experimental approaches there appeared to be roughly equal instances of attenuation and read-through, which was thought to be the result of differing riboswitch conformations. However, this alone was insufficient evidence of conformational switching, a defining characteristic of riboswitch function.

5.5 Riboswitch conformational equilibrium likely affects transcription attenuation

Structural modelling of the riboswitch predicts two possible conformations: a more thermodynamically favourable intrinsic terminator conformation and an anti-terminator or read-through conformation (see section 1.9). Riboswitches often exist in an equilibrium between a thermodynamically favourable default conformation and a less favourable conformation permissive through ligand interaction (Roth & Breaker 2009). However, transcriptional riboswitches may rely on ligand binding kinetics to dictate regulatory output, rather than conformational thermodynamics, especially if the equilibrium is slow to equate (Wickiser *et al.* 2005). The regulatory decision must be made before the RNA polymerase (RNAP) reaches the transcriptional expression platform. But by manipulating the thermodynamic stability of the riboswitch, it might be possible to assay riboswitch-mediated attenuation *in vivo* and potentially facilitate candidate ligand interaction.

5.5.1 Identification of potential mutation sites within the riboswitch sequence

Riboswitches are highly specialised regulatory RNAs with nucleic acid sequences that have evolved to allow more than one functional conformation (Blount & Breaker 2006). Accordingly, basepair interactions in any given conformation may vary greatly between conformations. For example, basepairing will be different in a terminator conformation versus the alternative read-through conformation. This makes it challenging to mutate a riboswitch sequence whilst maintaining functionality as mutations to one conformation must not have an unintended effect on the other, otherwise an affected conformation would require a compensatory mutation, which would in turn likely affect the previous conformation.

The *rpfB* riboswitch was scrutinised for residues capable of manipulating the free energy of one conformation with negligible impact to the alternative conformation. A limited number of sites were identified that could be exploited to strengthen basepairing within the stem regions of the conformations. Many of these sites were disregarded as the substitutions would affect loop regions in the alternative conformation, which could impact functionality as many riboswitches use complex ligand-loop or loop-loop interactions (Tucker & Breaker

2005; Wickiser *et al.* 2005; Montange & Batey 2008; Serganov 2010). Exhaustive inspection identified mutations capable achieving the desired impact on predicted folding thermodynamics with minimal impact to the riboswitch conformations.

First, a mutation capable of increased transcription attenuation and decreased read-through was devised. During intrinsic termination hairpin nucleation occurs, forming the terminator stem-loop which affects transcription termination (Peters, Vangeloff & Landick 2011; Ray-Soni, Bellecourt & Landick 2016). In theory, increased hairpin sequence complementarity could strengthen the secondary structure of the proposed intrinsic terminator, increasing the likelihood of transcription attenuation events. Inspection of the riboswitch terminator stem identified non-complementary basepairing between the adjacent G94:G112 residues (see Figure 5.VI). Modelling predicted a G112C substitution could strengthen terminator stem complementarity to lower stable folded free energy for the terminator conformation (-54.80 kcal/mol, -4.60 kcal/mol below the WT terminator conformation). Moreover, the G112C substitution destabilised the C2:G112 basepair within the anti-terminator conformation with structural models predicting no possible anti-terminator conformation.

For the anti-terminator or read-through conformation, the U6:G107 adjacent residues (see Figure 5.VI) within the initial 5'stem of the anti-terminator conformation were identified as a potential mutation site. Modelling predicted a U6C substitution could increase complementarity within the stem resulting in a marginally decreased free energy (-46.40 kcal/mol, -1.70 kcal/mol below the WT anti-terminator conformation). Additionally U6C substitution disrupted the U6:A51 basepair within terminator conformation resulting in an increased free energy (-48.30 kcal/mol, +1.90 kcal/mol above the WT terminator conformation). Although modelling still predicted the terminator conformation to be more thermodynamically favourable for stable folding, the predicted initial folding free energy (initial folding of a linear RNA molecule into a secondary structure) was marginally lower for the anti-terminator conformation (-47.40 kcal/mol, -1.1 kcal/mol below the U6C terminator at -46.10 kcal/mol). It was therefore decided the U6C substitution might be sufficient to observe an effect on the conformational equilibrium when compared to the WT riboswitch.

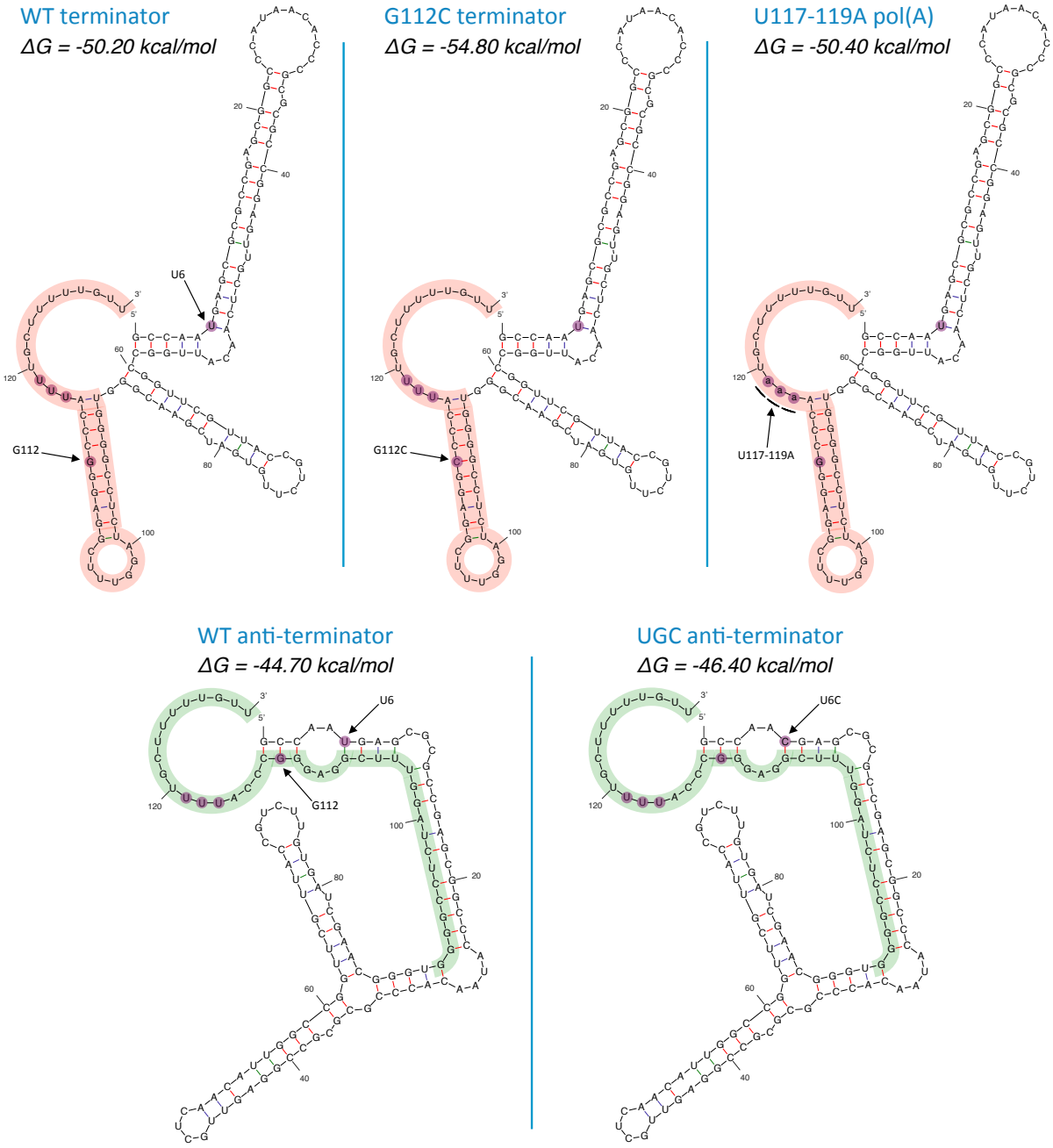


Figure 5.VI riboswitch conformational mutations
Structural modelling predictions of riboswitch sequence (Zuker 2003). WT terminator and anti-terminator structures are shown on the left, with identical red and green sequences highlighted for comparison. G112C terminator shows the terminator structure with stabilising mutation. U6C anti-terminator shows the anti-terminator structure with terminator destabilising mutation. U117-119A shows the terminator structure with poly(U) tract disrupting mutation. Mutation sites highlighted in purple for comparison.

A further mutation was devised to disrupt the poly(U) tract of the potential intrinsic terminator to explore the mechanism of attenuation. Transcriptional pausing within the poly(U) tract facilitates intrinsic termination by providing time for the terminator stem to fold and affect the elongation complex (Ray-Soni, Bellecourt & Landick 2016). The *rpfB*

riboswitch poly(U) tract is not a perfect poly(U) tract, but an interrupted stretch of uracil residues (UUUUGCUUUUGUU) not uncommon for mycobacterial intrinsic terminators (Czyz *et al.* 2014). It was hypothesised that increased transcriptional read-through could potentially be achieved through further interruption of this poly(U) tract. This would interfere with intrinsic termination either preventing the duration with which RNAP paused at the site or distancing the pause site from the forming terminator stem to prevent disruption to the elongation complex. Therefore the U117-119A substitution was identified as a suitable mutation (see Figure 5.VI), with structural modelling showing no predicted effect to either riboswitch conformation and a negligible effect on the predicted free energy (+0.20 kcal/mol increase relative to the WT terminator conformation).

5.5.2 Riboswitch mutants demonstrate a shift in conformational equilibrium

A riboswitch translational reporter fusion was made (identical to that employed for mapping of the translation start site (see section 3.2), containing WT *M. tuberculosis* H37Rv *rpfB* 5'UTR spanning -316 bp upstream of the re-annotated TTG start codon down, with the first 14 codons of *rpfB* CDS fused with *lacZ* (see Figure 5.VII, A). The construct was mutated individually at U6C, G112C and U117-119A relative to the P_1 TSS. Constructs were transformed into *M. smegmatis* as a surrogate host, and expression was measured by β-galactosidase assay.

Protein extracts of *M. smegmatis* transformed with WT translational fusion showed a β-galactosidase activity of 57.05 ± 2.27 Miller units (MU)/mg total protein, representative of WT riboswitch read-through (see Figure 5.VII, B). The terminator stabilising G112C substitution showed a near 2-fold significant decrease in expression to 30.33 ± 4.79 MU/mg total protein relative to the WT. Conversely the anti-terminator stabilising U6C substitution showed a significant increase in expression to 76.38 ± 10.39 MU/mg total protein relative to the WT. A similar result was seen with the terminator disrupting U117-119A substitution where expression increased significantly to 84.05 ± 2.94 MU/mg total protein relative to the WT. Taken together these results demonstrated the conformational mutations had the predicted effect as terminator conformation stabilisation increased attenuation but anti-terminator conformation stabilisation increased read-through. To confirm these results

were due to variation in riboswitch attenuation itself, total RNA was extracted from transformed *M. smegmatis* for northern blot analysis.

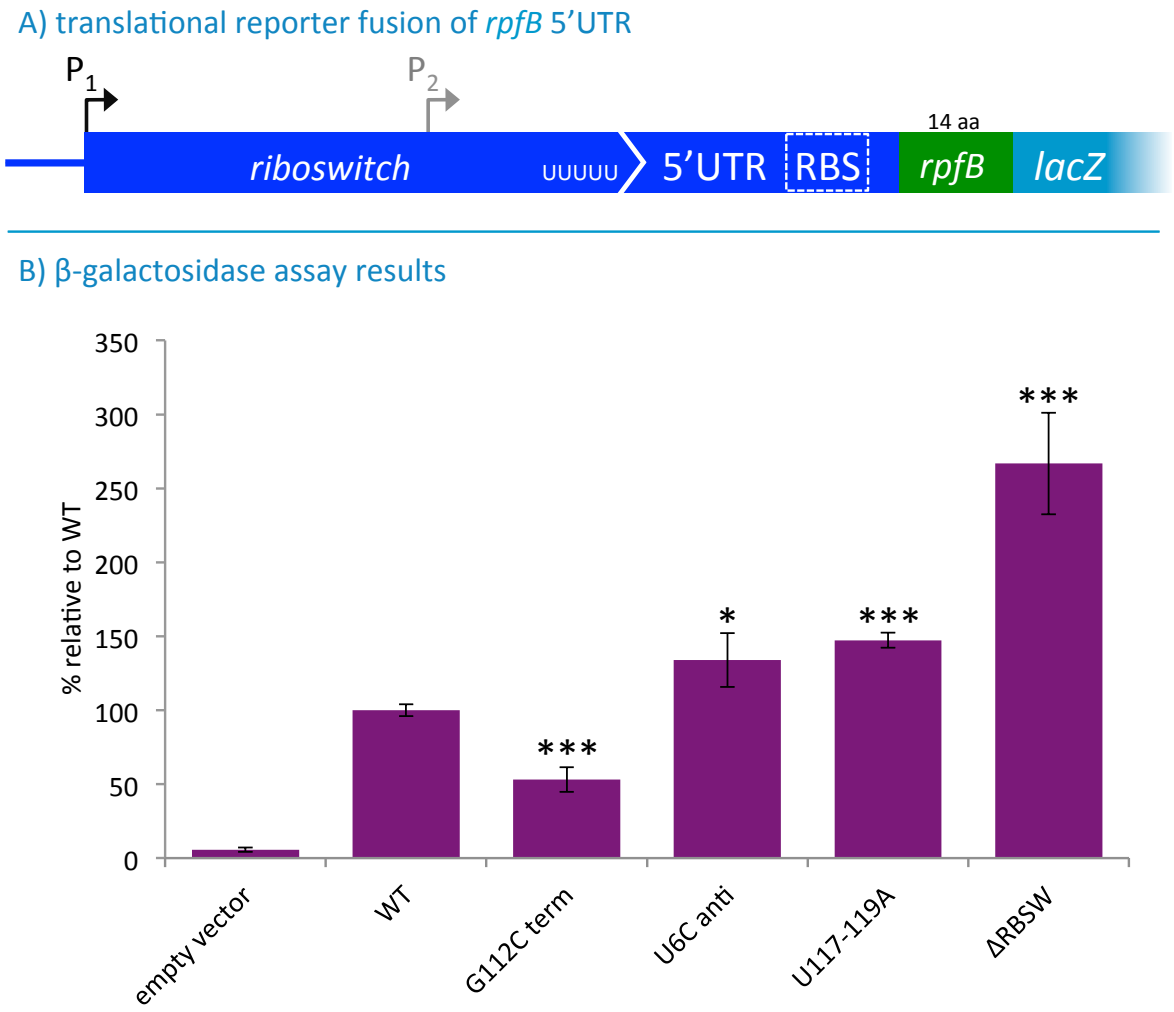


Figure 5.VII β-galactosidase assay of riboswitch conformational mutants
Diagram (A) shows the construction of the translational fusion employed. Chart (B) shows β-galactosidase assay results of transformed *M. smegmatis* protein extracts. Mean of biological replicates (n=6) relative to WT. Statistical significance relative to WT calculated by one-way ANOVA with Tukey post hoc analysis in IBM SPSS ±1 standard deviation, * $P \leq 0.05$, *** $P \leq 0.001$.

The resulting northern blot in Figure 5.VIII showed the same ~130 nt riboswitch attenuated transcript was present in the transformed *M. smegmatis,* WT *M. tuberculosis* and *Mycobacterium bovis* BCG. Evidently *M. smegmatis* did not express the riboswitch from the native *rpfB* P_1 promoter to the same extent as the other two species. This may be attributed to a lack of genetic context, as it has been shown that two-component response regulator MtrA and nucleoid associating protein Lsr2 are capable of binding within the *M. tuberculosis rpfB* locus, both of which have homologous *M. smegmatis* sequences but unverified function (Lew *et al.* 2011; Gordon 2013; Sharma *et al.* 2015).

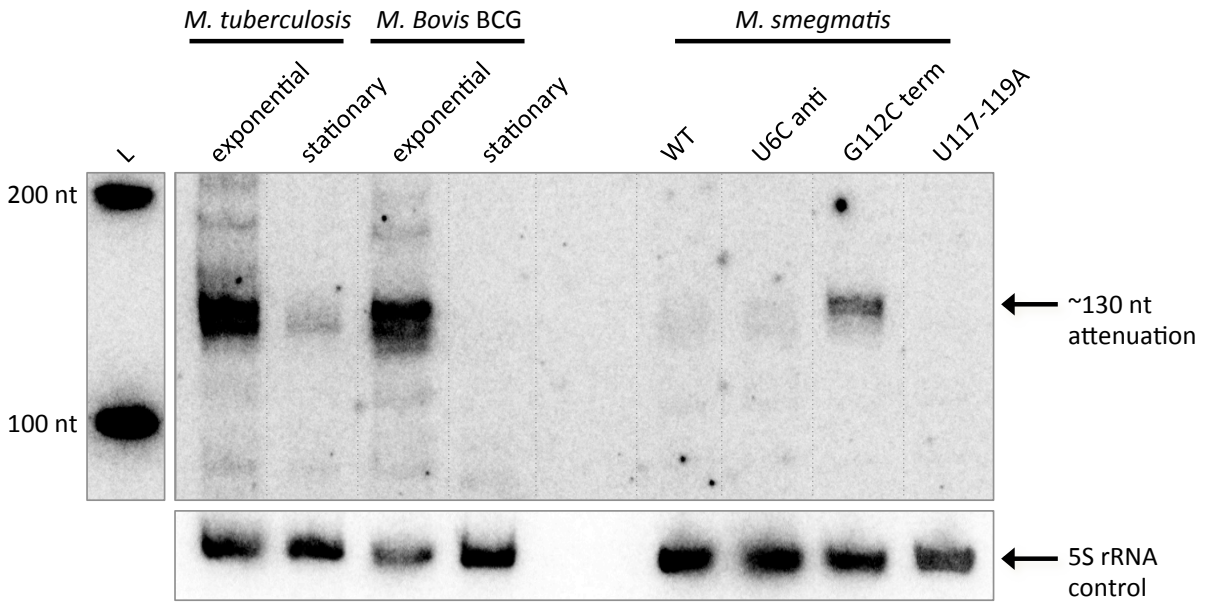


Figure 5.VIII northern blot of riboswitch reporters
Total RNA extracted from *M. tuberculosis* H37Rv and *M. bovis* BCG during both exponential and stationary growth phases on the left, on the right total RNA extracted during exponential growth phase from *M. smegmatis* transformed with riboswitch translational fusion reporter constructs containing U6C anti-terminator stabilising substitution, G112C terminator stabilising substitution and U117-119A terminator disrupting substitution. RNA was probed within the *rpfB* P_1 5'UTR (oligonucleotide 5.22) and for 5S rRNA as a loading control (oligonucleotide 1.48, see Appendix section 10.3).

Corresponding with decreased expression in the β-galactosidase assay above (Figure 5.VII), *M. smegmatis* extracts with the terminator stabilising G112C substitution resulted in increased abundance of the ~130 nt riboswitch attenuated transcript compared to WT. Moreover, the terminator disrupting U117-119A substitution had no detectable attenuated transcript, likely due to decreased efficiency of intrinsic termination at the disrupted poly(U) tract. Collectively the northern blot of the *M. smegmatis* reporter strains and the β-galactosidase assay results further substantiate the proposition that the riboswitch regulates expression of downstream genes through transcription attenuation, likely achieved by intrinsic termination at the poly(U) tract.

5.5.3 Riboswitch deletion demonstrates a strong downstream transcriptional control

A further construct was devised to ascertain the magnitude of regulation on downstream gene expression. The WT reporter described above was mutated, deleting the riboswitch sequence from +1 to +127 bp relative to the P_1 TSS (this deletion was made prior to 3'RACE analysis (see section 5.4) and G128-U130 remained in the fusion). Riboswitch deletion resulted in a near 3-fold significant increase in β-galactosidase expression to

152 ± 19.58 MU/mg total protein relative to the WT (see Figure 5.VII). This large increase in expression indicated that the riboswitch exerts a strong control on transcription of downstream genes, principally the *rpfB-ksgA* bi-cistron in the context of *M. tuberculosis* (see section 4.2).

5.6 Chapter discussion

Northern blot analysis of total RNA from *M. tuberculosis* and *M. bovis* BCG demonstrated a distinct lack of riboswitch complementary sequence during stationary phase growth (Figure 5.VIII). This was not surprising as resuscitation promoting factors serve to stimulate growth during favourable conditions, thus expression during stationary growth phase would likely be disadvantageous to cells. Decreased expression of *rpfB* into early and late stationary growth phase has been described in the literature previously, it is therefore unlikely *rpfB* is expressed as cells continue into late stationary growth phase (Tufariello, Jacobs & Chan 2004; Gupta, Srivastava & Srivastava 2010).

Unlike *M. tuberculosis*, *M. bovis* BCG stationary growth phase RNA yielded a faint signal of ~130 nt attenuated transcript. As the two species share 100% sequence identity within the *rpfB* locus, this discrepancy cannot be attributed to the riboswitch itself. Additionally, the 5S rRNA loading control appeared to be consistent between the two species. In the experiments above, stationary growth phase RNA extraction was carried out 1 week after cultures achieved turbidity in excess of 1.0 $OD_{600\ nm}$. For the strains used in this study *M. bovis* BCG was observed to have a slightly longer generation time of 24 hours compared to that of *M. tuberculosis* at approximately 16 hours. Therefore the stationary growth phase cultures may not have been directly comparable between the two species; *M. bovis* BCG cultures may have been slightly earlier in stationary phase than *M. tuberculosis* at the time of sample extraction. It is probable this disparity may be associated with the phenotypic heterogeneity associated with entry into stationary growth phase. Regardless, it is clear both species demonstrated a marked reduction in transcription during stationary growth phase with little riboswitch attenuated transcript and no read-through transcripts detectable.

In contrast, the abundance of signal in exponential growth phase RNA on the analysed northern blot (Figure 5.VIII) supported the theory that the riboswitch expression regulates

resuscitation promoting factors under advantageous growth conditions. As the P_2 promoter is situated within the riboswitch, it is incapable of yielding a riboswitch attenuated transcript of ~130 nt. Reporter fusions (Figure 5.III) demonstrated expression is likely promoted by P_1 as P_2 appeared to make an unsubstantial contribution to expression when the P_1 promoter is compromised.

Notwithstanding, it was surprising to observe that disruption to the P_2 -10 element resulted in increased reporter expression compared with the WT. The cause of this increase is unknown. One theory is that disruption to the P_2 -10 element could have resulted in decreased RNAP association at the P_2 promoter. An alternative explanation is that the mutations made to the P_2 -10 element had unforeseen effects to the riboswitch itself in which the P_2 promoter is situated. Structural modelling did not predict any deviation to either of the riboswitch conformations or the folding free energy associated with the substitutions. However, the U73C substitution (relative to the P_1 TSS) used to change the downstream thiamine of the P_2 -10 element would have substituted a residue within a loop region of both riboswitch conformations which may have impacted functionality. In this instance, such a disruption was unavoidable, but its potential to contribute to this increased expression cannot be disregarded.

All of the reporters fusions employed within this study were initially constructed in *Escherichia coli*. Generally, *E. coli* is a poor host for mycobacterial genetics where promoters generally have a poor -35 promoter element consensus making sigma factor recognition difficult (Cortes *et al.* 2013). In preliminary experiments the promoter reporters were transformed into *M. smegmatis* for validation prior to experimentation in *M. tuberculosis*. Within an *M. smegmatis* background there was no detectable P_2 expression, but subsequent experiments proved limited P_2 expression occurred within an *M. tuberculosis* background (Figure 5.III). Consequently, *M. smegmatis* makes an excellent host in which to investigate the riboswitch function because the vast majority of transcripts made will encode the riboswitch promoted by the upstream P_1 promoter. Despite this, the question was raised as to what the biological function of the P_2 promoter was in the wider context of *rpfB* expression.

As experiments have shown the P_2 promoter has scarce activity in either exponential or stationary growth phase cultures of *M. tuberculosis*, it is possible that the P_2 promoted expression might be associated with physiological stress. The sequence of the P_2 -10 element bears close resemblance to the SigA -10 element 'TANNNT' consensus (Cortes *et al.* 2013). SigA is the primary sigma factor of *M. tuberculosis* and promotes expression during normal growth conditions (Newton-Foot & Gey Van Pittius 2013). A lack of P_2 promoted expression suggests the P_2 promoter may be subjected to additional regulation such as the two-component response regulator MtrA or the nucleoid associating protein Lsr2 which are known to associate with the region (Lew *et al.* 2011; Gordon 2013; Sharma *et al.* 2015). Perhaps under certain conditions it may be advantageous for cells to express the *rpfB-ksgA* bi-cistron separately from riboswitch regulatory control through P_2 promoted transcription. This is certainly a potential area for future investigation, although not pursued in the present study concerned with the riboswitch-mediated regulation.

Recently Sharma *et al.* (2015) described a SigB promoter -10 element upstream from a potential transcription start site for *rpfB* identified through 5'RACE. The 5'end identified was situated immediately downstream of the riboswitch described in this study at G128 (relative to the P_1 TSS) indicating the possibility of a processed 5'end. Sharma *et al.* investigated the SigB element through reporter assays using transformed *M. tuberculosis* cell extracts grown to mid-exponential growth phase. The SigB sigma factor is primarily associated with stresses connected to stationary phase growth in *M. tuberculosis* (Newton-Foot & Gey Van Pittius 2013). Assaying potential SigB activity in mid-exponential growth phase is therefore somewhat inappropriate. Moreover, Sharma *et al.* deduced there were no other possible promoters within the *rpfB* locus. They proceeded to test the *rpfB* locus using *in vitro* transcription assays consisting of *M. tuberculosis* core RNAP in the presence and absence of SigB using a template spanning +45 to +187 bp (relative to the P_1 TSS), excluding the P_1 promoter described in the present study. The transcription reaction generated a single transcript, which was not sized with a molecular marker and therefore is not sufficient evidence to determine functionality of the proposed SigB promoter.

In the present study two promoters have been validated, both of which are corroborated by existing TSSM data published by Cortes *et al.* (2013) prior to the study by Sharma *et al.* (2015). By digesting with terminal-exonuclease (TEX), an RNase capable of degrading

truncated 5'monophosphorylated RNA, Cortes *et al.* enriched their samples for primary 5'triphosphorylated transcripts prior to sequencing. The 5'RACE method Sharma *et al.* employed did not differentiate between primary 5'transcripts and processed 5'ends, which are relatively more abundant. For this reason it is likely the TSS identified by Sharma *et al.* is an artefact of post-transcriptional processing rather than a true TSS. In the present study, exhaustive efforts were made to carry out 5'RACE through a wide variety of 5'RACE methods. Unfortunately, none of the techniques used successfully identified a 5'terminus to the *rpfB* 5'UTR. Potentially the low expression and highly structured nature of the riboswitch may have confounded 5'adapter ligation and reverse transcriptase efficiency. Regardless, the results from the P_1 and P_2 double mutant (Figure 5.III) indicated there were no other regions capable of promoting transcription in the *rpfB* locus, in agreement with TSSM published by Cortes *et al.* (2013).

In addition to verifying the sense strand promoters proposed for the *rpfB* locus, transcriptional fusion of the putative P_{as} promoter verified it is capable of promoting transcription *in vivo*. Figure 5.III shows there was a discernible level of β-galactosidase reporter expression from the transcriptional fusion of the P_{as} promoter. Considering the construction of this transcriptional reporter differs from that of the sense strand promoter translational reporters, it would be inappropriate to compare promoter strength based on reporter expression alone because the experimental set up is not comparable (Thomas 2001). It is likely the asRNA transcript is not highly expressed as efforts to map the asRNA by 3'RACE or detect its expression by northern blot analysis repeatedly proved unsuccessful throughout this study.

Taken in consideration with the weak signal observed in RNA-seq data (Figure 5.I) it is likely that the asRNA transcript is less abundant than *rpfB* sense strand transcripts. It is possible this may be attributed to comparably weaker promoter activity, transcript stability and turnover or potentially the function of the antisense transcript itself which may serve to basepair with the *rpfB* sense transcript as an RNA:RNA duplex susceptible to RNase degradation (Wagner, Altuvia & Romby 2002). Potentially the asRNA may have greater relevance under certain stress conditions. In a similar manner to the P_2 promoter, the P_{as} promoter resembles the SigA -10 element 'TANNNT' consensus sequence, but does not appear to be highly expressed during exponential growth. As discussed in previous chapters

(see section 3.4), it is feasible the asRNA transcript may down-regulate *rpfB* expression through occlusion of the ribosome binding site. This could make a promising facet of future investigation with the potential to elucidate further complexity in *rpfB* expression regulation.

The 3'temini of riboswitch attenuated transcripts were mapped to the poly(U) tract of the riboswitch intrinsic terminator through 3'RACE. Of the clones sequenced 42.42% mapped specifically to the poly(U) tract itself (Figure 5.IV), consistent with transcription attenuation in this region either through intrinsic termination or long-lived transcriptional pausing. Interestingly, a relatively high 7.58% of clones sequenced mapped to G179 (relative to the P_1 TSS), the last nucleotide of the re-annotated TTG start codon (see section 3.2). Recently, a thiamine phosphate (TPP) sensing riboswitch was described in *E. coli* that uses a translation initiation proximal pause as an "intramolecular lock", a failsafe mechanism to ensure the regulatory decision riboswitch expression platform is adhered to (Chauvier *et al.* 2017). It is tempting to speculate such a secondary function may be achieved by the *rpfB* riboswitch where results indicated the possibility of start codon proximal pausing (see Appendix Figure 10.IV).

With regards to the mechanism by which the riboswitch exerts its regulatory control, the conformational reporter mutants (Figure 5.VII) strongly indicated the riboswitch functions as an intrinsic terminator. Upon stabilising the predicted terminator conformation of the riboswitch with the G112C substitution, there was a significant decrease in transcriptional read-through as indicated by decreased β-galactosidase expression. Through northern blot analysis of transformed *M. smegmatis* total RNA extracts, a marked increase in ~130 nt riboswitch attenuated transcript was detected compared to the WT reporter, indicative of increased attenuation. Combined with the U117-119A terminator disrupting mutation, which displayed a significant increase in transcriptional read-through, the results strongly indicate the *rpfB* riboswitch uses the proposed intrinsic terminator to affect transcriptional attenuation of downstream genes.

It is encouraging that the both the terminator stabilising G112C substitution and the anti-terminator stabilising U6C substitution resulted in significant deviations from WT reporter expression, fitting trends predicted by the structural modelling (Figure 5.VII). The

absence of any empirical structural observations of the riboswitch is a noteworthy limitation to the present study. Although extensive endeavours were made to attempt to probe the structure of the riboswitch, isolation of highly purified, homogenously sized riboswitch transcript proved elusive to generate. Despite this it is probable the predicted *rpfB* riboswitch terminator conformation represents a functional transcription attenuating structure, as demonstrated by 3'RACE (Figure 5.IV), conformational reporter mutants (Figure 5.VII) and northern blot (Figure 5.VIII). Although, it is possible the anti-terminator or read-through conformation could be quite different in the presence of bound ligand. Nevertheless, the two predicted *rpfB* riboswitch conformations demonstrate a hallmark of riboswitch function, a mutually exclusive switch where formation of one conformation precludes that of the other (Batey 2011; Millman *et al.* 2017).

Figure 5.IX demonstrates how the stem-loop region of the intrinsic terminator is mutually exclusive to the terminator conformation, precluding formation of the predicted anti-terminator conformation and *vice versa*. This is further ratified by the observation that both the G112C and U6C substitutions strengthen the basepairing of one conformation whilst simultaneously weakening the other (Figure 5.VI). This attribute made both of these substitutions elegant solutions to investigate the potential conformational equilibrium of the riboswitch without compromising the overall structure. In future chapters these mutations will be validated yet further by demonstrating their effect on riboswitch-mediated attenuation *in vitro* (see section 7.2.4).

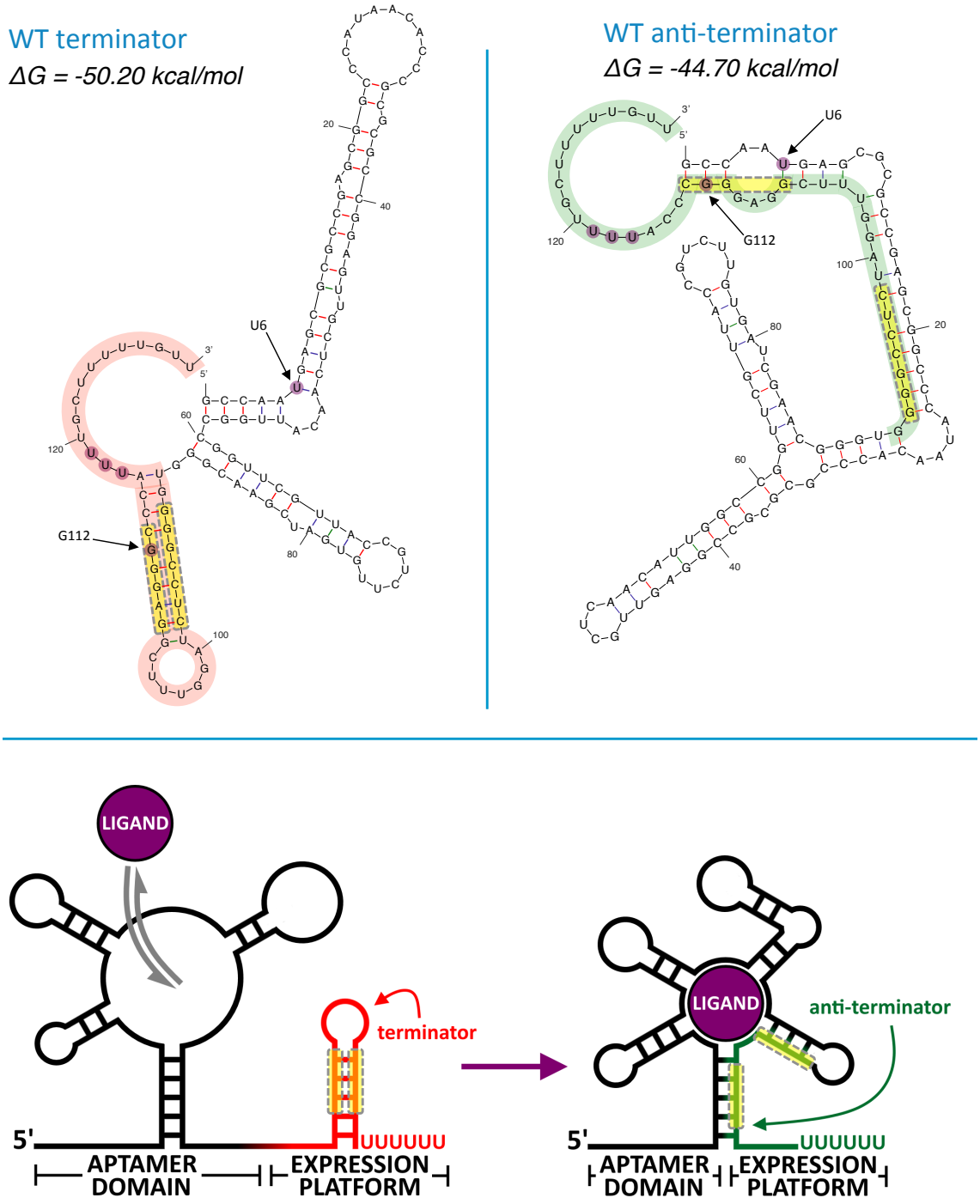


Figure 5.IX a mutually exclusive switching mechanism
The top diagrams show the predicted *rpfB* riboswitch conformations with the stem region of the intrinsic terminator stem-loop highlighted in yellow (Zuker 2003). The bottom diagram demonstrates how one conformation precludes the formation of another in a riboswitch mutually exclusive switch (adapted from Batey 2011).

5.7 Chapter conclusion

In conclusion, the results described in this chapter provided substantial evidence that the proposed *rpfB* riboswitch exerts regulatory control on downstream genes through transcription attenuation, likely achieved by intrinsic termination or a very long-lived transcriptional pause. The ability for the riboswitch to switch between its terminating conformation to a mutually exclusive alternative conformation permissive of transcription read-through was demonstrated through minimal perturbation. In addition, the putative P_1, P_2 and P_{as} promoters were experimentally validated, corroborating published transcription start site mapping and indicating the potential for an extra facet of *rpfB* regulation in the form of an antisense sRNA.

6 *In vivo* riboswitch-associated expression

6.1 Diversity of growth environments

The work described in previous chapters provides insight into expression of the *rpfB* locus throughout exponential and stationary phase growth. However, these growth states are laboratory models, controlled environments in which microorganisms may be investigated. In natural environments, *Mycobacterium tuberculosis* experiences a range of physiological stresses throughout transmission, infection, immune evasion, dormancy and resuscitation. Consequently, the diverse environmental conditions and stimuli require multiple phenotypic sub-populations with varied gene expression.

The work detailed in this chapter aimed to further understanding of how the *rpfB* locus is regulated under varying growth conditions, and to what extent expression is controlled by the riboswitch itself.

6.2 Rapid riboswitch expression upon exiting stationary growth phase

As demonstrated in previous chapters (see section 5.2) the riboswitch is not expressed during stationary phase growth. However, RpfB functions to facilitate resuscitation from dormancy under advantageous conditions. It was therefore hypothesised it might be

possible to detect riboswitch expression as cells resumed exponential growth by re-suspending late stationary growth phase culture in fresh media. Accordingly, *M. tuberculosis* H37Rv was grown to late stationary phase (two weeks after 1.0 $OD_{600\,nm}$) before being re-suspended in fresh media. Total RNA was extracted from the re-suspended culture as a time-course from 0 to 7.5 hours, size separated by polyacrylamide gel electrophoresis (PAGE) and blotted onto a nylon membrane to be probed.

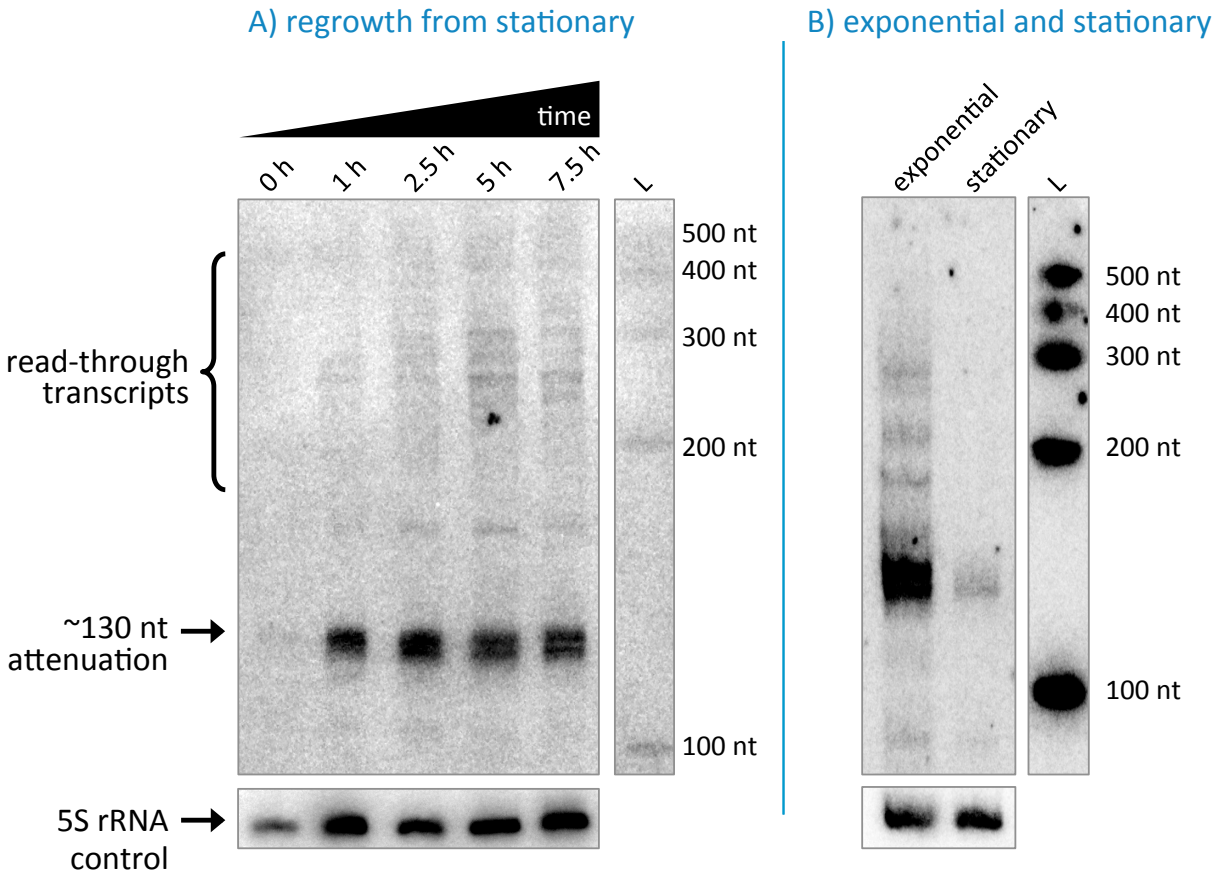


Figure 6.I northern blot re-growth from stationary growth phase
Image (A) shows northern blot of total RNA extracted from *M. tuberculosis* H37Rv during re-growth from stationary growth phase. Late stationary growth phase culture was re-suspended in fresh media (see section 2.2.2). Total RNA was extracted from samples as a time-course from 0 hours following re-suspension. ~130 nt attenuation represents riboswitch-mediated attenuation (see section 5.2). Image (B) is a replica of Figure 5.II for comparison of the ratio of attenuated transcript. RNA was probed within the *rpfB* P_1 5'UTR (oligonucleotide 5.22) and for 5S rRNA as a loading control (oligonucleotide 1.48, see Appendix section 10.3).

The resulting northern blot (see Figure 6.I, A) showed riboswitch expression was resumed within 1 hour of cells being re-suspended in fresh media, as represented by visualisation of the ~130 nt riboswitch attenuated transcript (see section 5.2). In contrast to northern blots of exponential growth phase total RNA (see Figure 6.I, B), there was substantially more

attenuated transcript compared to read-through. This suggested that read-through beyond the terminator did not occur immediately as cells resumed exponential growth, as there appeared to be relatively more terminated transcripts.

6.3 Riboswitch transcripts are not expressed during nutrient starvation

A previous study by Gupta, Srivastava & Srivastava (2010) investigated expression of *M. tuberculosis rpf*(*A-E*) under multiple physiological stresses and growth conditions. Expression was measured by real-time quantitative PCR (RT-qPCR), using cDNA reverse transcribed from extracted mRNA. Values were normalised to 16S ribosomal RNA (rRNA). In addition to exponential and stationary growth phases, samples were collected during nutrient starvation, acid stress, and hypoxia as well as non-culturable and resuscitation cultures. Gupta, Srivastava & Srivastava (2010) reported the largest change in *rpfB* expression occurred within cultures subjected to 24 hour nutrient starvation, where a 19-fold increase was reported. To investigate if increased *rpfB* expression could be the result of increased riboswitch read-through, a nutrient starvation experiment was conducted in the present study.

Exponential growth phase *M. tuberculosis* culture was pelleted and washed in phosphate buffered saline (PBS) with 0.05% Tween80 (v/v final) twice before being re-suspended in an equal volume of PBS Tween80 (see section 2.2.3) (Gengenbacher *et al.* 2010; Gupta, Srivastava & Srivastava 2010). Total RNA was extracted from the re-suspended culture as a time-course from 0-24 hours, size separated by PAGE and blotted onto a nylon membrane to be probed.

Figure 6.II shows riboswitch-associated expression of *rpfB* was substantially reduced during nutrient starvation. There was a rapid decrease in both read-through and attenuated transcript in later time-points compared to samples taken immediately after re-suspension. After 24 hours there was only a marginal restoration of ~130 nt riboswitch attenuated transcript, although the 5S rRNA loading control was also comparatively weaker. Nevertheless, this data indicated any potential increase in *rpfB* expression as reported by Gupta, Srivastava & Srivastava (2010) was unlikely to be promoted by the *rpfB* P_1 promoter or regulated by the riboswitch.

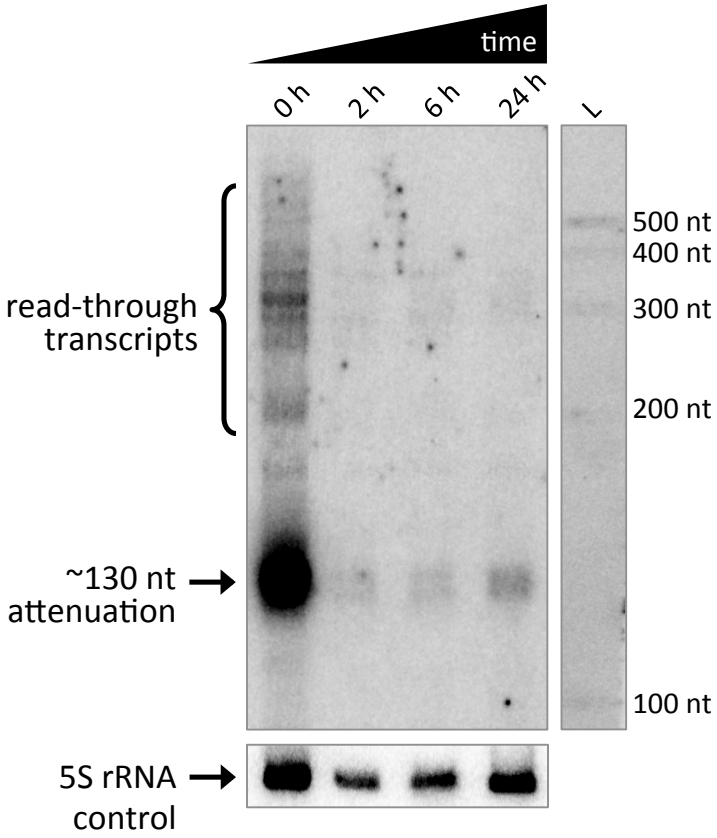


Figure 6.II northern blot of nutrient starvation
Northern blot of total RNA extracted from *M. tuberculosis* H37Rv during nutrient starvation. Exponential growth phase culture was re-suspend in PBS + 0.05% Tween80 for nutrient starvation (see section 2.2.3). Total RNA was extracted from samples as a time-course from 0 hours following re-suspension. ~130 nt attenuation represents riboswitch-mediated attenuation (see section 5.2). RNA was probed within the *rpfB* P_1 5'UTR (oligonucleotide 5.22) and for 5S rRNA as a loading control (oligonucleotide 1.48, see Appendix section 10.3).

6.4 Riboswitch transcripts are subject to rapid turnover

For any actively growing cell, transcript abundance *in vivo* is representative of transcript turnover, a balance between transcript synthesis and degradation. Transcripts that are detectable through molecular techniques always have a higher rate of synthesis than degradation, but if transcripts are particularly stable, with structures less accessible to RNase degradation, this can lead to transcript accumulation (Moores *et al.* 2017). To investigate if the abundance of riboswitch attenuated transcript observed through northern blot analysis was representative of low transcript turnover, a transcription inhibition time-course experiment was conducted.

M. tuberculosis culture was grown to exponential phase before adding rifampicin to inhibit transcription (see section 2.2.3). An untreated culture was grown in parallel and total RNA was sampled over 24 hours for northern blot analysis.

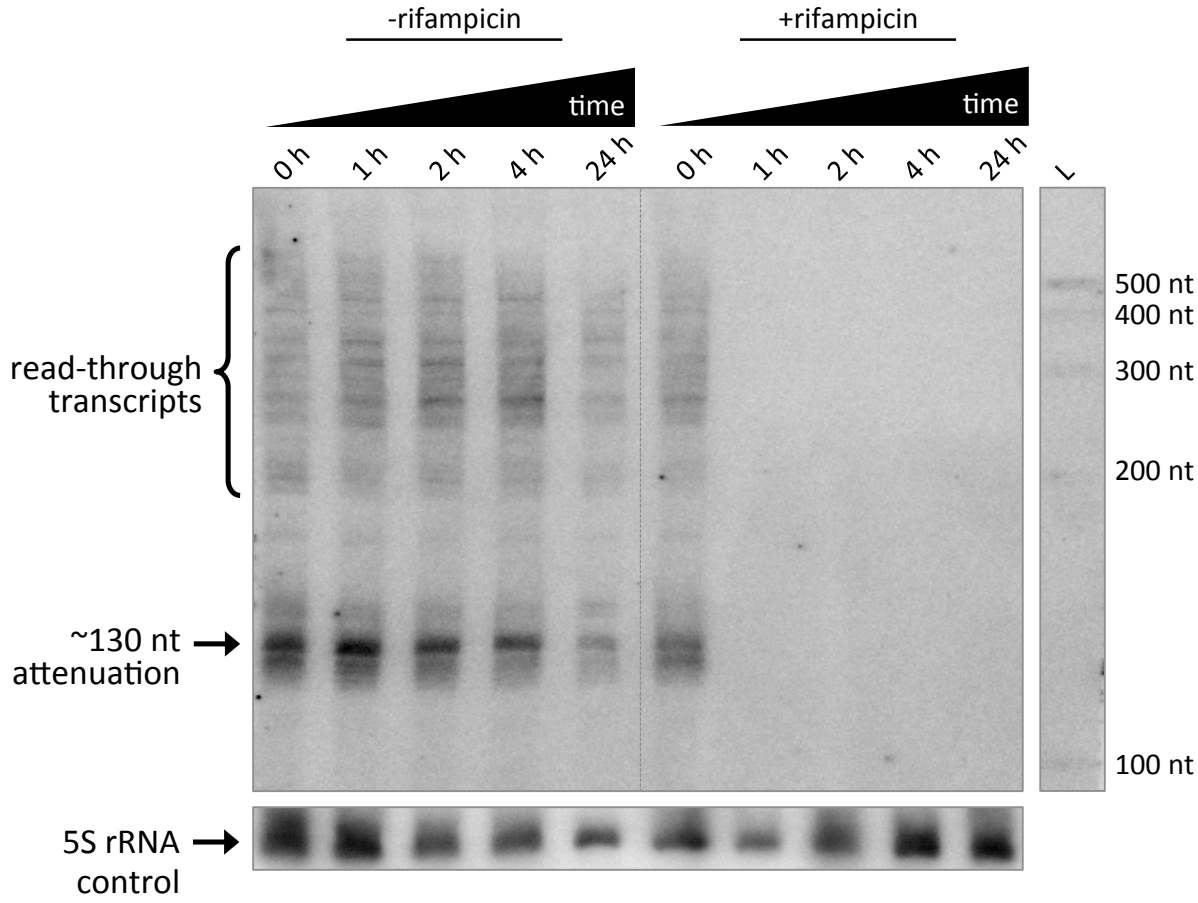


Figure 6.III northern blot of transcription inhibition
Northern blot of total RNA extracted from *M. tuberculosis* H37Rv during transcription inhibition. Total RNA was extracted from samples as a time-course from 0 hours ± 200 µg/mL rifampicin (see section 2.2.3). ~130 nt attenuation represents riboswitch-mediated attenuation (see section 5.2). RNA was probed within the *rpfB* P_1 5'UTR (oligonucleotide 5.22) and for 5S rRNA as a loading control (oligonucleotide 1.48, see Appendix section 10.3).

Within 1 hour of addition of rifampicin (Figure 6.III, right) there was total loss of any detectable signal for ~130 nt attenuated transcripts or read-through. This indicated transcript abundance was not the result of low transcript decay but rather continual transcript synthesis. In the absence of rifampicin expression remains constant, albeit with a marginal increase of read-through transcripts at 2 and 4 hour time-points (Figure 6.III, left). Unlike some small RNAs (sRNAs) previously reported (Moores *et al.* 2017), the riboswitch attenuated transcript likely decays quickly. In future, a more accurate measure of riboswitch transcript decay rate would require a more discrete time-course within 1 hour.

6.5 Expression of the *rpfB* locus throughout biofilm formation

Bacterial biofilms represent a heterogeneous population of growth phenotypes, a state distinct from planktonic growth or growth on solid agar media (Ojha *et al.* 2008). Mycobacterial biofilms utilise secreted mycolic acids to create a structured extracellular microenvironment that requires extensive changes in gene expression to form and colonise (Ojha *et al.* 2005; Richards & Ojha 2014); modulation of cell wall biosynthesis has been shown to be particularly important for biofilm formation in *Mycobacterium smegmatis*. However, in pathogenic mycobacteria, the *rpfB* riboswitch regulates the *rpfB-ksgA* bi-cistron, controlling expression of both the *rpfB* cell wall hydrolase and *ksgA* methyltransferase involved in ribosome biogenesis. It was therefore hypothesised that riboswitch-mediated regulation might be relevant to expression of either *rpfB* or *ksgA* during biofilm formation.

To explore this hypothesis, biofilms were cultured using *Mycobacterium bovis* BCG, a tractable model of *M. tuberculosis* that shares 100% sequence conservation for the *rpfB* locus. Biofilm formation was carried out in 50 mL polypropylene tubes, inoculating 10 mL of 7H9 media without the addition of Tween80, which would have disrupted aggregation (see section 2.2.4). Cultures were inoculated with an *M. bovis* BCG starter culture to a final density of 0.2 $OD_{600\,nm}$ before allowing biofilm formation under static non-aerated culture conditions. Biofilms formed as pellicles at the air-liquid interface, developing within the first 3 weeks. Pellicle layers were harvested from separate cultures over the course of 12 weeks. Total RNA was extracted from pellicle layers to generate cDNA through reverse transcription, which was used as template for RT-qPCR. Reactions were primed for multiple targets within the *rpfB* locus: riboswitch and riboswitch-mediated read-through, asRNA, *rpfB*, and *ksgA* (see section 2.7.6).

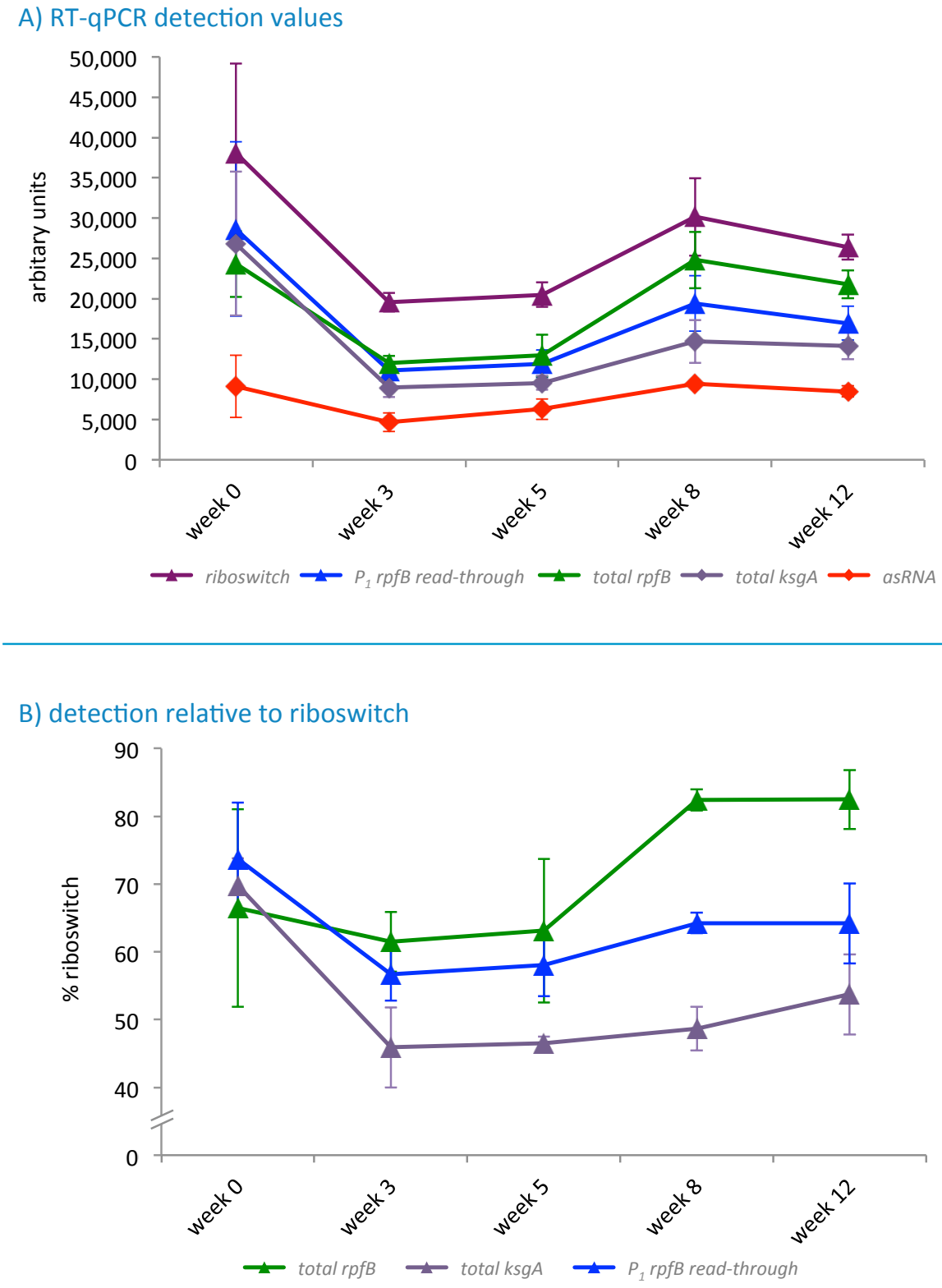


Figure 6.IV detection values of *M. bovis* BCG biofilm RT-qPCR
Chart (A) shows detected amplicons of individual *M. bovis* BCG biofilm cDNA targets. Values are expressed as arbitrary units calculated against a gDNA standard curve. Chart (B) shows detection relative to riboswitch amplicons (see section 2.7.6). Data represents mean of technical replicates (n=4) of the biological replicates (n=3) ± standard deviation.

There was roughly a 50% decrease in the expression of all target amplicons during initial biofilm formation from 0 to 3 weeks post-inoculation (see Figure 6.IV, A). However, expression of all targets increased during subsequent weeks as biofilms continued to mature. Intriguingly the variation between biological replicates across all target amplicons appeared to decrease as biofilms matured. This represents conditions that favour a biofilm forming phenotype, with increased variation of later time-points being indicative established sub-populations within the matured biofilm (Richards & Ojha 2014).

Amplicons generated by priming reactions within the riboswitch itself may represent either riboswitch-attenuated transcripts or read-through transcripts, as both contain the riboswitch sequence (see section 2.7.6). By priming reactions forward within the riboswitch and reverse within *rpfB* coding sequences (CDS), the 'P_1 *rpfB* read-through' amplicon was theoretically representative of transcripts that were not subject to riboswitch-mediated attenuation but continued read-through into *rpfB*. Therefore expressing 'P_1 *rpfB* read-through' results as a percentage of riboswitch transcripts provided a measure of riboswitch attenuation of P_1 promoted transcripts.

Prior to biofilm formation at week 0, P_1 *rpfB* read-through amplicons were detected at 73.68% ± 8.35 relative to detected riboswitch transcripts (see Figure 6.IV, B). After 3 weeks of static growth, thin pellicle layers were observed in cultures, where RT-qPCR indicated P_1 *rpfB* read-through decreased to 56.65% ± 3.84. However, by 8 weeks, mature biofilms exhibited P_1 *rpfB* read-through at 64.25% ± 1.51. Conversely, results for amplicons generated by priming alone within the *rpfB* CDS indicated total *rpfB* expression at 82.42% ± 1.56 in week 8 biofilms. Transcripts of the *rpfB* CDS may originate from either of the P_1 or P_2 transcription start sites (TSS) upstream, but only P_1 promoted transcripts would be subjected riboswitch regulatory control. As total *rpfB* expression appeared to exceed that of P_1 *rpfB* read-through, these results suggested that the P_2 promoter made an increasing contribution to the expression of *rpfB* as biofilms matured.

Whilst total *ksgA* expression appeared to follow a similar trend to that of total *rpfB* expression, unlike other targets *ksgA* expression did not resume the same level in mature biofilms. In fact, although not statistically significant, a declining trend in total *ksgA* expression relative to total *rpfB* expression was observed (see Appendix Figure 10.V). The

mechanism through which relative ratios of *rpfB* and *ksgA* might vary individually within the *rpfB-ksgA* bi-cistron is unknown. Potentially, the nutrient depleted media may have caused a reduction in available nucleotides that would have decreased RNA polymerase processivity. This would explain decreased transcription of *ksgA* as it is further downstream from the promoter than *rpfB*.

6.6 Chapter discussion

Upon resuming exponential growth, northern blot analysis of late stationary growth phase cultures exhibited rapid expression of attenuated transcripts within 1 hour of re-suspension in fresh media (Figure 6.I, A). Comparatively, read-through transcripts were less expressed until approximately 5 hours after re-suspension. Further biological replicates would be required to confirm this is not a stochastic effect, especially as the 5S rRNA loading control indicated some variation between time-points. Despite this, delayed expression of read-through transcripts could be indicative of riboswitch function as an intrinsic terminator. As a transcriptional 'on-switch', it is predicted ligand interaction would enhance read-through beyond the *rpfB* riboswitch (see section 1.9). Riboswitches can be sensitive to ligand abundance; it has been shown some may only switch at specific ligand concentration thresholds (Lutz *et al.* 2014). In the early stages of transition from stationary to exponential phase growth, cells may have had insufficient concentrations of cognate ligand to permit conformational switching and subsequent transcriptional read-through. Therefore it may be that the ligand is present during exponential growth phase, but read-through is only achieved at a concentration threshold.

The sequence homology of the *M. tuberculosis rpfB* riboswitch appears to be conserved in pathogenic mycobacteria only (see section 1.9). As *M. tuberculosis* is an obligate pathogen, it is plausible the cognate ligand may be associated with pathogenesis, perhaps a host derived signal of infection or weakened immunity used to indicate conditions advantageous to resuscitation by RpfB. Whilst it would be intriguing to investigate riboswitch-mediated expression of the *rpfB-ksgA* bi-cistron within an infection model such as macrophage tissue culture, results of the work described in this chapter suggested cognate ligand may not be associated with pathogenicity.

Limited riboswitch expression was detected during nutrient starvation conditions. Northern blot analysis of exponential growth phase cultures re-suspended in PBS demonstrated a rapid down-regulation of both riboswitch attenuation and read-through transcripts (Figure 6.II). Riboswitch expression was marginally resumed within 24 hours of nutrient starvation. This suggests the 19-fold increase in *rpfB* expression observed by Gupta, Srivastava & Srivastava (2010) was unlikely the result of riboswitch-mediated regulation. Moreover, as both riboswitch attenuated and read-through transcripts require P_1 promoted transcription, it is also unlikely this increased *rpfB* expression would have been promoted by the P_1 promoter.

The ribo-probe used to visualise transcripts by northern blot analysis (Figure 6.II) was complementary to the riboswitch, capable of binding transcripts upstream of the P_2 promoter TSS and therefore unable to bind P_2 promoted transcripts. As described in previous chapters, the P_2 promoter does not appear to make any substantial contribution *rpfB* expression during exponential growth or growth on solid agar media (see section 5.3). P_2 promoted transcription of the *rpfB* locus independent of riboswitch-mediated regulatory control could be more relevant under starvation conditions. This theory would not only explain the increased *rpfB* expression described by Gupta, Srivastava & Srivastava (2010) during nutrient starvation, but would corroborate increased expression of total *rpfB* CDS observed in mature *M. bovis* BCG biofilms where cells would likely also experience nutrient depletion (Figure 6.IV, B). In future, it would be worthwhile investigating expression of both P_1 promoted riboswitch transcripts as well as P_2 transcripts using RT-qPCR. Additional physiological stresses could also be explored such as osmotic stress similar to the *rpfA ydaO* riboswitch homologue (St-Onge *et al.* 2015), acid stress like the pH sensing ribo-regulator (Nechooshtan *et al.* 2009; Dambach *et al.* 2015), and resuscitation from dormancy by supplementation of cell-free exponential growth phase culture supernatant (Mukamolova, Turapov, Kazarian, *et al.* 2002; Gupta, Srivastava & Srivastava 2010) to further understand expression and regulation of the *rpfB* locus.

Northern blot analysis of cultures subjected to transcription inhibition (Figure 6.III) demonstrated riboswitch transcripts were turned over within 1 hour of inhibition. A potential cause of rapid turnover of riboswitch transcripts may be interaction with antisense transcripts promoted by P_{as}. Attempts to characterise expression of the putative asRNA by

3'rapid amplification of cDNA ends or northern blot analysis proved unsuccessful in this study. Reporter fusions of the asRNA promoter region indicated low promoter activity on solid agar media (see section 5.3). Despite this, it was possible to detect asRNA target amplicons by RT-qPCR of biofilm extracts described above (Figure 6.IV, B). Results indicated the putative asRNA to be the least expressed of all amplified targets, consistently approximately 4-fold lower than detected riboswitch expression. Potentially this is indicative of asRNA function, where antisense transcripts might basepair with riboswitch sense transcripts to form RNA:RNA duplexes as substrate for RNase degradation. Alternatively, antisense P_{as} promoter activity may simply serve to reduce template availability for the P_1 and P_2 sense strand promoters, as the convergent promoters would cause polymerase clashing.

Multicellular structures have been described in caseous granulomas of mycobacterial infections (Canetti 1965, reviewed in Ojha & Hatfull 2012) and many mycobacteria, including *M. tuberculosis*, readily form biofilms in static culture (Zambrano & Kolter 2005; Ojha *et al.* 2008). Multiple models have been proposed for the cultivation of biofilms as pellicles, solid attachments or flow cells (Branda *et al.* 2005). Pellicle formation provides a three dimensional structured microenvironment where multiple phenotypic sub-types may grow. The RT-qPCR results obtained from *M. bovis* BCG biofilms in the present study demonstrated large variability resulting in statistical insignificance for trends observed in expression of amplified targets (Figure 6.IV). Variability in samples extracted at week 0 may have been the result of heterogeneity of the starter culture inoculum. For later time-points, variability may have been due to the heterogeneity associated with established biofilms (Richards & Ojha 2014). It is also possible the precision of the sample extraction process (see section 2.2.4) may have contributed to this variation[1]. Independent of these considerations, RT-qPCR technical replicates demonstrated excellent reproducibility within each biological replicate. In future, to substantiate *rpfB* locus expression during biofilm formation with statistical significances would require a greater sample size of biological replicates.

[1] Personal communication, Dr Jim Huggett, LGC UK National Measurement Laboratory/University of Surrey, 2017.

6.7 Chapter conclusion

In conclusion, the work described in this chapter demonstrated cultures transitioning from late stationary phase into exponential growth exhibit delayed riboswitch-mediated transcription read-through, indicative of potential ligand concentration dependent riboswitch conformational switching. Contrastingly, during nutrient starvation and biofilm formation *rpfB* expression appeared to be independent of riboswitch expression, potentially promoted by the *rpfB* P_2 promoter. Finally, transcription inhibition indicated that riboswitch transcripts are rapidly turned over.

7 Optimisation of a riboswitch *in vitro* transcription system

7.1 Investigating riboswitch function *in vitro*

The work described in previous chapters investigated the mechanism through which the *rpfB* riboswitch exerts its regulatory control *in vivo*. The culmination of these findings strongly indicated the riboswitch operates on a transcriptional level, regulating expression of downstream genes through transcriptional attenuation. The riboswitch encodes a potential intrinsic terminator, a stem-loop followed by a downstream poly(U) tract. Many riboswitches exist in a thermodynamic equilibrium between two possible conformations, however the regulatory output of transcriptional riboswitches are often dictated by ligand binding kinetics rather than conformational thermodynamics (Wickiser *et al.* 2005). It has been demonstrated that conformational switching is likely achieved within discrete ligand concentration thresholds (Lutz *et al.* 2014). Many riboswitches respond to critical metabolites such as S-adenosyl-L-methionine (SAM) or nucleotide derivatives (Smith *et al.* 2010; St-Onge *et al.* 2015) therefore perturbation of transcriptional riboswitches *in vivo* is particularly challenging as cells have multiple mechanisms of regulating intracellular concentrations of such metabolites. *In vivo* systems are further compounded by culture heterogeneity and extended growth times particularly for mycobacteria. *In vitro*

transcription systems offer a controlled environment in which the concentration of candidate ligand can be varied to investigate possible effects on riboswitch function.

The work described in this chapter details the development and optimisation of an *in vitro* transcription system capable of exploring the riboswitch transcription attenuation regulatory mechanism.

7.2 Optimisation of an *in vitro* transcription system using *E. coli* RNAP

7.2.1 Development of a riboswitch *in vitro* transcription template

The first step in the development of the *in vitro* transcription system was to design a suitable transcription template. For the purpose of developing and optimising the *in vitro* transcription system it was decided commercially produced *Escherichia coli* RNA polymerase (RNAP) would be preferable as a consistent and highly purified source of RNAP for the assay. Therefore a promoter capable of transcription in both *E. coli* and mycobacteria would be ideal. Preliminary experiments *in vivo* had demonstrated the native *rpfB* P_1 promoter had relatively week activity, which was poorly recognised in *E. coli* due to the absence of any defined -35 promoter element. Previously, the P_1 promoter of the *Mycobacterium smegmatis* ribosomal RNA *rrnB* operon was identified as a highly active promoter capable of recognition by RNAP in *E. coli* and mycobacterial species due to the presence of both -10 and -35 promoter elements, making it a promising candidate for the *in vitro* transcription system (Arnvig *et al.* 2005).

An *M. smegmatis rrnB* P_1 promoter (P_{rrnB}) fragment spanning -80 to -1 bp of the P_{rrnB} transcription start site (TSS) was fused with +1 to +176 of the riboswitch (relative to the *rpfB* P_1 TSS). The entire construction was cloned into the sub-cloning vector pGEM (Promega). Initially the template was isolated and purified by restriction digestion, however this resulted in an abundance of non-specific transcripts, likely due to the presence of cleavage overhangs (Artsimovitch & Santangelo 2015). Subsequently templates were generated by PCR amplification. Preliminary experiments indicated the P_{rrnB} promoter was poorly recognised by *E. coli* holoenzyme as non-specific 'end-to-end' transcription of the template occurred frequently (see Appendix Figure 10.VI). Consequently the promoter was modified to include an extended -10 element with the 'TGN' motif (Agarwal & Tyagi 2003; Ruff,

Record & Artsimovitch 2015) substituting 'ACGTAACTT' with 'tggTAACTT' to enhance promoter activity (Arnvig 2001). The extended -10 P_{rrnB} template demonstrated significantly reduced non-specific transcription and increased promoter specificity (see Appendix Figure 10.VII). Moreover, a substantial proportion of transcripts generated were approximately ~130 nt long, indicating probable riboswitch-mediated attenuation as this size corresponded with the sequence of the riboswitch. The presence of a large discrete signal also indicated transcription read-through as template run-off.

It was hypothesised that addition of candidate ligand could affect a shift in signal from attenuation to read-through as run-off if the ligand were recognised by the riboswitch. Following the discovery of the *rpfB-ksgA* bi-cistron described in previous chapters (see section 4.2) it was thought the riboswitch might be more closely associated with regulation of the methyltransferase *ksgA* than *rpfB.* Therefore SAM, the essential co-factor of KsgA, was selected as a rational candidate ligand for initial screening. However, reactions with SAM indicated ligand addition had a deleterious effect on transcription initiation[2], with transcription substantially decreased across all concentrations of SAM tested (see Appendix Figure 10.VIII). Therefore a system was developed where candidate ligand could be added to reactions after initiation.

[2] Personal communication, Dr Wade C. Winkler, University of Maryland (2016).

7.2.2 Single-round transcription of the *rpfB* riboswitch

Withholding any given rNTP from a transcription reaction allows the formation of halted transcription elongation complexes (TECs) that stall upon encountering the corresponding template base for which rNTP is unavailable. The ability to halt TECs allows multiple aspects of *in vitro* transcription reactions to be modulated without affecting initiation (Landick, Wang & Chan 1996; Grundy, Winkler & Henkin 2002; Sudarsan *et al.* 2003; Wickiser *et al.* 2005):

- Radiolabelled nucleotide can be utilised during initiation and subsequently out-competed with high concentrations of non-labelled nucleotide during elongation, effectively end-labelling transcripts. Uniform radiolabelling allows accurate detection of different sized transcripts, rather than biasing detection towards longer transcripts, which have more radiolabelled nucleotide incorporated.
- High concentrations of heparin can be added to the halted TECs without affecting initiation. Any dissociated RNAP will be unable to re-initiate, limiting reactions to single-round transcription. As a result, transcripts are representative of single transcription events.
- The concentration of the full complement of rNTPs can be varied to affect TEC processivity as elongation resumes. Moreover, transcription is resumed as a synchronised event, allowing samples to be taken chronologically as a time-course.
- Transcription can be initiated with high concentrations of dinucleotide corresponding to +1 and +2 positions of the template TSS. This allows substantially increased initiation efficiency as the dinucleotide increases stability of initiating RNAP and specificity to the desired TSS (Phillips *et al.* 2001). Low concentrations of rNTPs during initiation also serve to reduce read-through beyond the stall site by decreasing TEC processivity (Artsimovitch & Santangelo 2015).
- Reactions can be initiated as a batch so that halted TEC aliquots can be used individually to reduce variability between reactions. Thus, candidate ligand or transcription factors can be added to the reactions and allowed to interact without affecting initiation before resuming elongation.

The formation of stable halted TECs requires at least 10 nt of initially transcribed sequence (Landick, Wang & Chan 1996; Artsimovitch & Henkin 2009). As the riboswitch sequence utilises all four bases within the first 6 nt, the template required modification to make it suitable for the formation of halted TECs. Through careful scrutiny, the addition of a 5 nt 'GTAAA' sequence to the 5'end of the riboswitch sequence was devised to increase initially transcribed sequence length. The addition of 5'GTAAA showed no change to the predicted structural modelling (Zuker 2003) of either riboswitch conformation (see Figure 7.I).

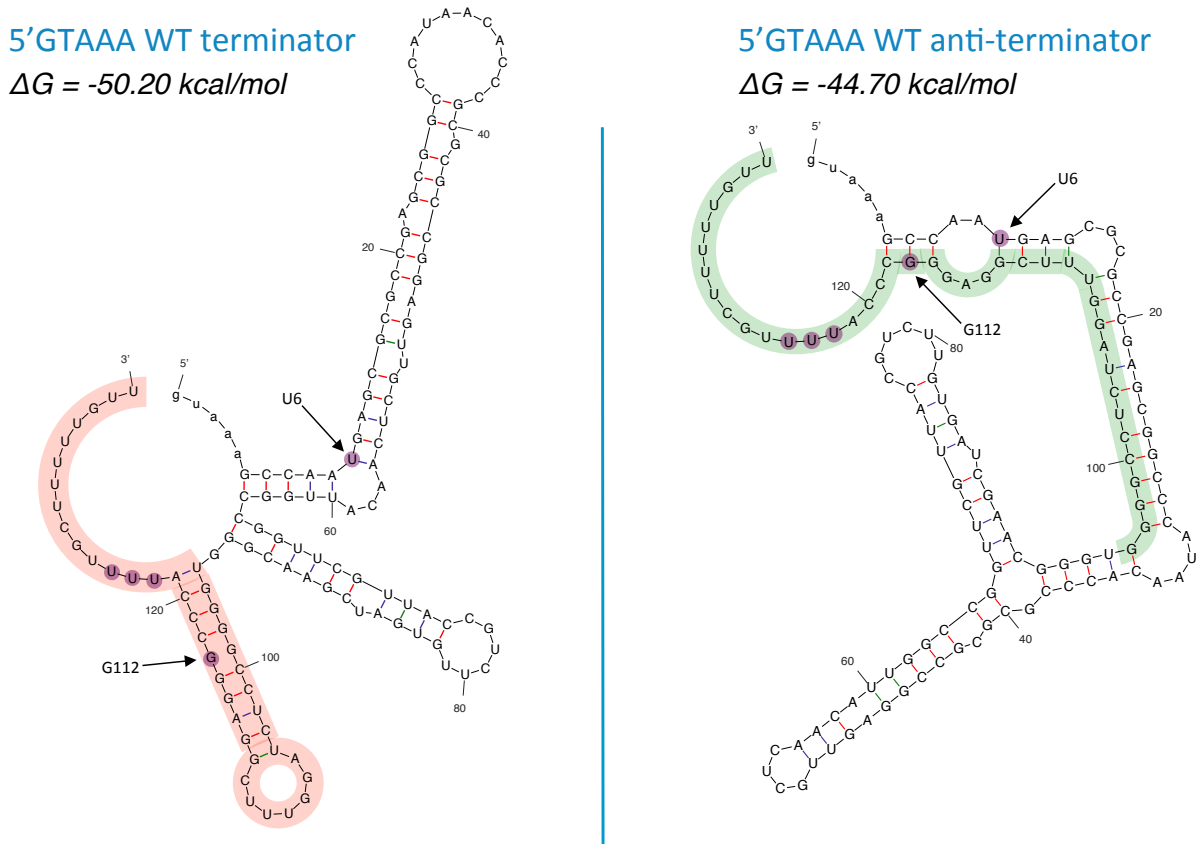


Figure 7.I 5'modified riboswitch conformations
Predicted structural modelling of the riboswitch modified for the formation of halted transcription elongation complexes with the addition of 'GTAAA' to the 5'terminus. Identical red and green sequences highlighted for comparison. Mutation sites highlights in purple for comparison.

With the addition of the 5'GTAAA sequence to the template, UTP could be withheld from reactions initiated with GpU dinucleotide to halt transcription at U+11 (U+6 relative to *rpfB* P_1 TSS). Whilst this would yield a very short initially transcribed sequence, more substantial modification would only further increase probability of affecting the riboswitch structure or function. Therefore the 5'GTAAA sequence was added to the template between the P_{rrnB} TSS and riboswitch 5'terminus. The template was subsequently used to generate halted TECs in an *in vitro* transcription reaction as described in (see section 2.9), withholding UTP during initiation and radiolabelling with 32P-αATP. With the 5'GTAAA additional sequence the revised template was theoretically capable of riboswitch-mediated attenuation at ~130 nt or template run-off at 191 nt.

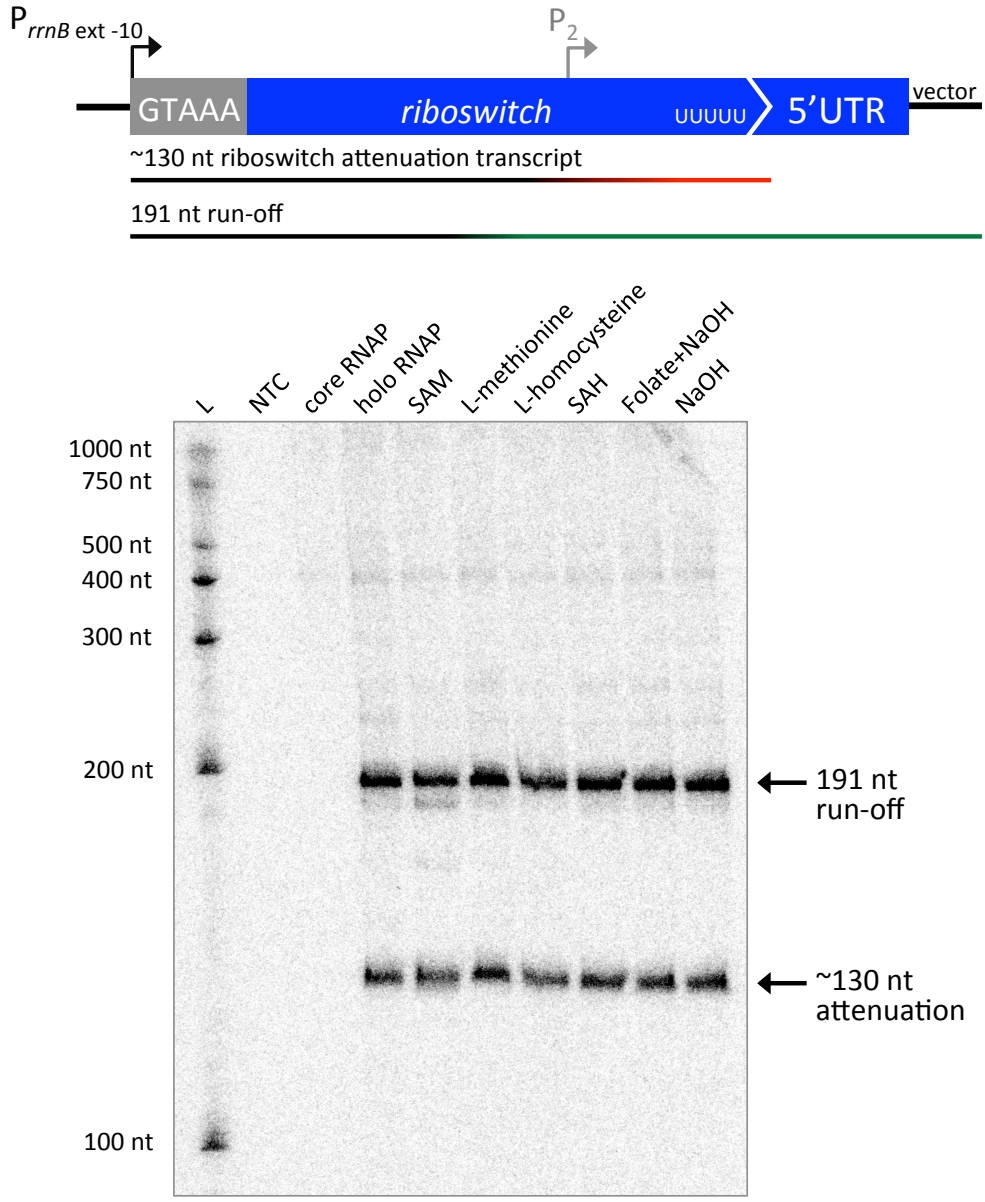


Figure 7.II *in vitro* transcription with ligand candidates
Single-round *in vitro* transcription of the riboswitch using core *E. coli* RNAP and holoenzyme with candidate ligand. From left to right: 300 µM S-adenosyl-L-methionine (SAM), 200 µM L-methionine, 200 µM L-homocysteine, 200 µM S-adenosyl-L-homocysteine (SAH), 200 µM folate in 1 M NaOH, and 1 M NaOH control. Reactions were carried out at 37°C with 17 nM *E. coli* RNAP (NEB), initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Elongation was resumed with 10 µM rNTPs, incubating for 20 min before stopping reactions with the addition of formamide loading dye. Template amplified with oligonucleotides 5.31 and 5.23 (see Appendix section 10.3).

To confirm promoter specificity, the template was tested with *E. coli* core RNAP (NEB). Figure 7.II shows neither the ~130 nt attenuation or the 191 nt run-off transcripts were generated by core RNAP. Comparatively, the *E. coli* holoenzyme (core RNAP with *E. coli* σ^{70}) produced discrete signals for both transcripts, confirming promoter specificity. Moreover,

there was a distinct absence of non-specific transcripts, most likely achieved by dinucleotide initiation and addition of high concentrations of heparin (200 ng/µL) following initiation.

Unfortunately neither SAM nor any associated methyl cycle intermediate (Parveen & Cornell 2011) candidate ligand tested resulted any variation to the ratio of attenuation:read-through. Although SAM did not have any effect on attenuation, there appeared to be transcripts specific to the addition of SAM. This was further investigated and found to be due to an additional 10 nt of vector sequence on the 3'terminus of the template included to increase the size of run-off transcripts. Subsequently the template was amplified without the additional vector sequence, which abrogated these transcripts (data not shown). The revised 261 bp template (shown in Figure 7.III) demonstrated riboswitch-mediated attenuation at ~130 nt and transcription read-through as template run-off at 181 nt.

7.2.3 Addition of NusA reveals long-lived pauses during riboswitch transcription

Co-transcriptional ligand interaction can be very important for transcriptional riboswitches (Wickiser *et al.* 2005), with transcriptional pausing allowing correct folding of riboswitch domains (Nechooshtan *et al.* 2009; Frieda & Block 2012; Perdrizet *et al.* 2012). Discrete pausing can also favour formation of one riboswitch conformation over the other, even if less thermodynamically stable (Steinert *et al.* 2017). In many instances, transcriptional pausing is mediated by transcription factors, principally NusA which has been shown to increase the frequency and duration of pausing (Artsimovitch & Landick 2000).

In the presence of NusA reduced RNAP processivity could facilitate nascent RNA folding and ligand binding during *in vitro* transcription of the *rpfB* riboswitch (Ha *et al.* 2010). Preliminary experiments indicated an optimum molar ratio of 1:5 *E. coli* holoenzyme to recombinant *E. coli* NusA (NEB) respectively (see Appendix Figure 10.IX). Subsequently the riboswitch was transcribed *in vitro* in a batch single-round reaction, adding NusA to halted TECs prior to elongation. Aliquots were taken periodically over a 20 minute elongation period and mixed with formamide loading dye to inhibit further transcription (see section 2.9).

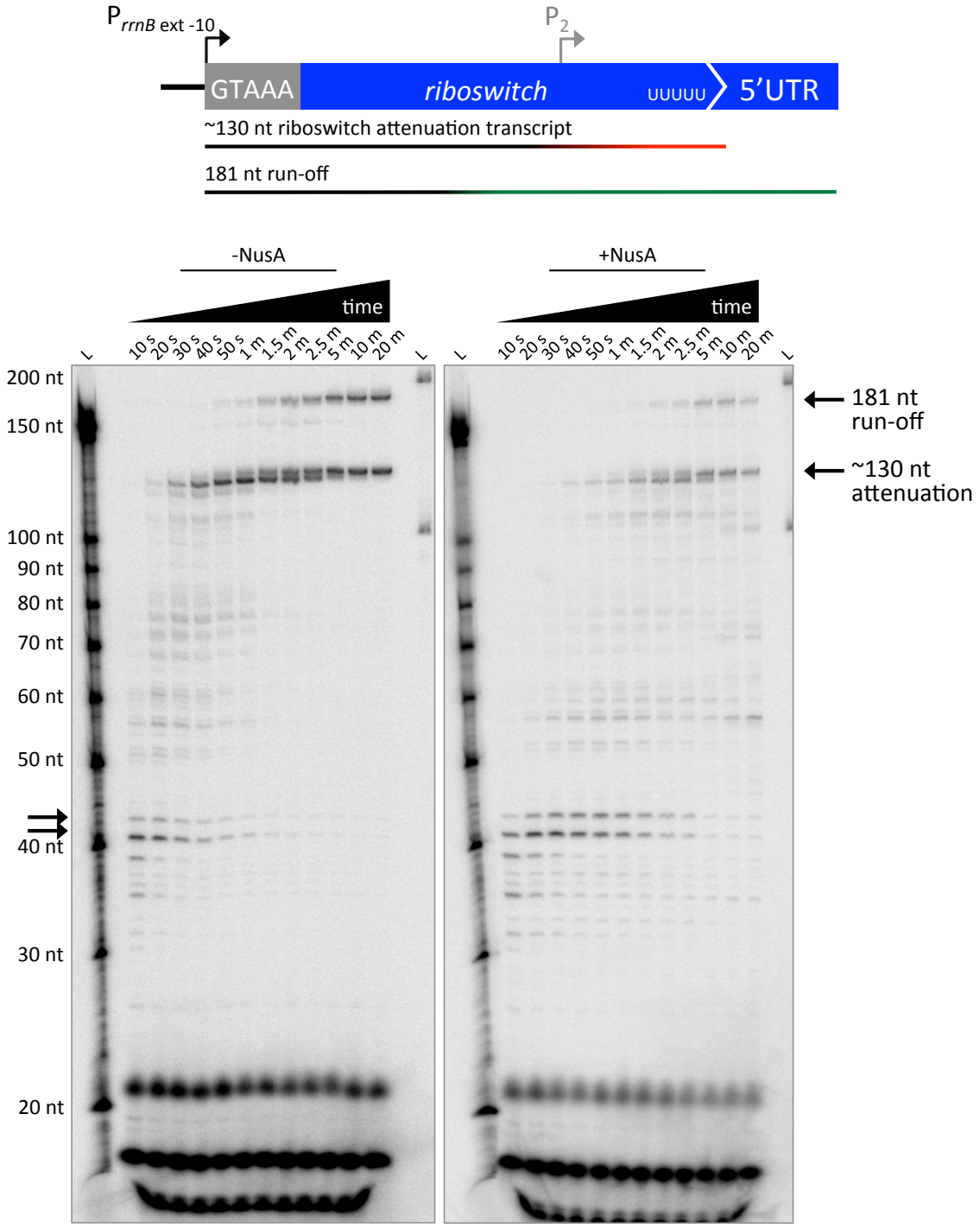


Figure 7.III *E. coli in vitro* transcription time-course
Single-round *in vitro* transcription time-course in the presence and absence of NusA. Reactions were carried out at 30°C to decrease processivity, with 17 nM *E. coli* RNAP (NEB), initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Reactions were carried out ± 85 nM NusA (NEB). Elongation was resumed with 25 µM NTPs, incubating for up to 20 min before stopping reactions with the addition of formamide loading dye. Template amplified with oligonucleotides 5.31 and 5.23 (see Appendix section 10.3).

Figure 7.III demonstrates the transcription of the riboswitch over time. Three intense bands indicative of halted TEC abortive transcripts were observed at the bottom of the gels. Abortive transcripts did not appear to resume elongation throughout the time-course. Above the abortive transcripts a series of discrete pauses occurred over the time-course. As expected, pausing was substantially enhanced in the presence of NusA. Long-lived pauses were identified at 41 and 43 nt (G+36 and G+38 less 5'GTAAA), which increased in duration 5-fold in the presence of NusA.

Two bands were visible at the site of the predicted terminator. The smaller of these bands decreased in abundance over time, resulting in increased abundance of both the ~130 nt attenuation and 181 nt run-off transcripts. This indicated terminator proximal pausing is required prior to attenuation or read-through, a pre-requisite of intrinsic terminators which require poly(U) tract pausing to facilitate hairpin formation (Zhang & Landick 2016). Collectively, the results indicated NusA might be required *in vivo* to enhance pausing and hence facilitate folding of riboswitch intermediates within a timeframe permissive of cognate ligand interaction.

7.2.4 Riboswitch mutants demonstrate a shift in conformational equilibrium *in vitro*

Through single point mutations the riboswitch conformational equilibrium can be shifted in favour of transcription attenuation or read-through (see section 5.5). It was theorised that the same mutations could be made to the transcription template to explore the effect on riboswitch transcripts. Accordingly, mutations were made to the riboswitch sequence and cloned into the template to form halted TECs. Separate mutations consisted of the terminator conformation stabilising G112C (C+117 relative to the P_{rrnB} TSS), anti-terminator conformation stabilising U6C (C+11) and intrinsic terminator poly(U) tract disrupting U117-119A (U+122 to U+124). Structural modelling predicted no change to the structure or stable folding free energy with the required 5'GTAAA addition (Zuker 2003).

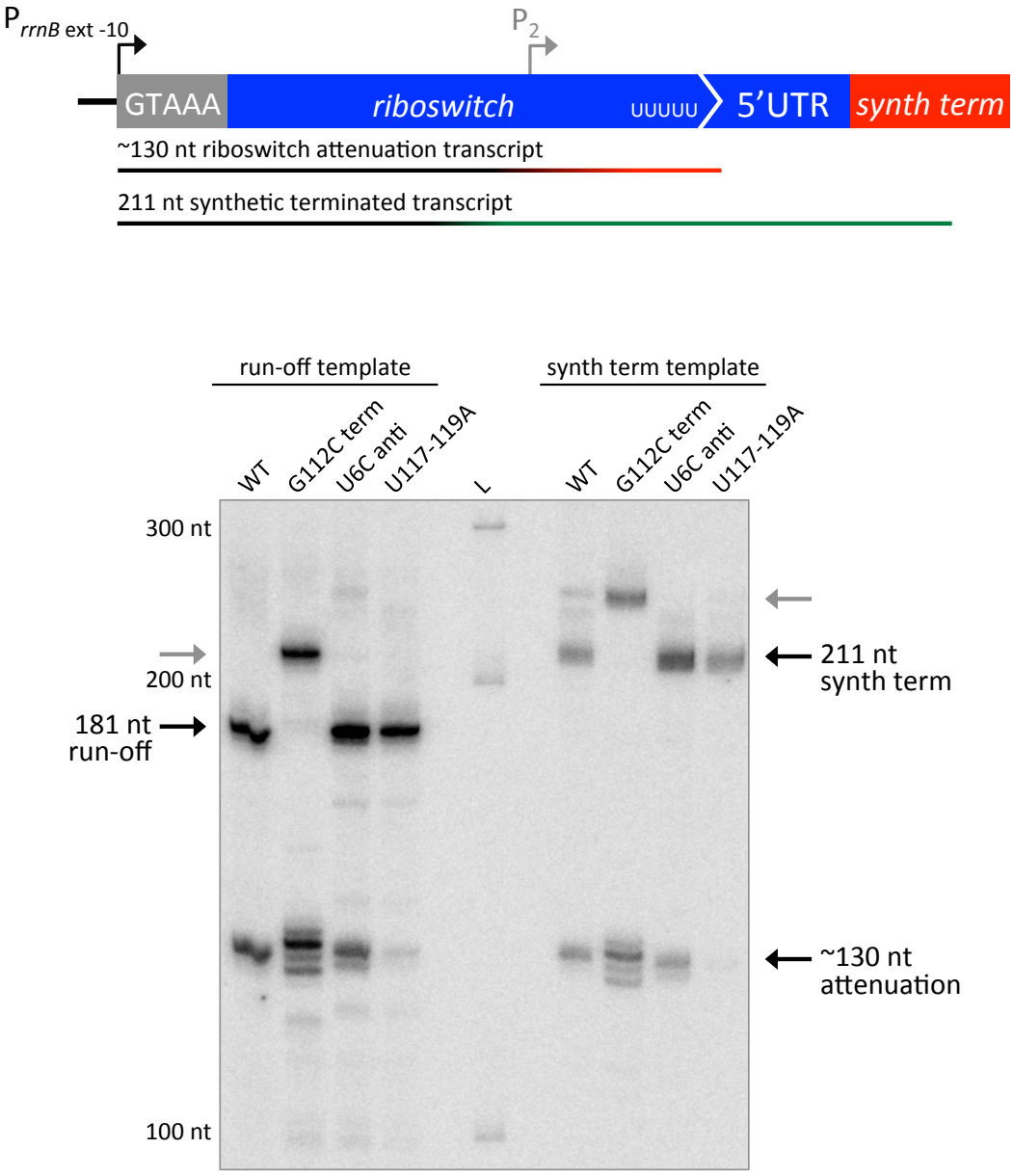


Figure 7.IV *E. coli in vitro* transcription of riboswitch mutants
Single-round in vitro transcription of riboswitch conformational mutants. Left) the WT riboswitch template, followed by the G112C terminator stabilising substitution, the U6C anti-terminator stabilising substitution and U117-119A intrinsic terminator disrupting substitution. Right) templates with additional *synB* synthetic terminator added downstream. Grey arrows denote G112C transcription artefact. Reactions were carried out at 30°C to decrease processivity, with 17 nM *E. coli* RNAP (NEB), initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Reactions were carried out in the presence of 85 nM NusA (NEB). Elongation was resumed with 25 µM rNTPs, incubating for 20 min before stopping reactions with the addition of formamide loading dye. Run-off template amplified with oligonucleotides 5.31 and 5.23 (see Appendix section 10.3). Synthetic terminator template amplified with oligonucleotides 5.44 and 8.43 (see Appendix section 10.3).

The mutant templates demonstrated definite variations in transcript abundance when compared to the WT (see Figure 7.IV, left). The poly(U) tract disrupting U117-119A substitutions indicated near complete disruption to riboswitch-mediated attenuation at ~130 nt. Similarly, the anti-terminator conformation stabilising U6C substitution resulted in decreased termination in favour of increased read-through, albeit to a lesser extent. Conversely, the intrinsic terminator stabilising G112C substitution resulted in substantially increased attenuation generating multiple discrete attenuated transcripts, thought to be the result of increased intrinsic termination efficiency. However, the G112C substitution also exhibited a read-through transcript much larger than any other template generated.

At approximately 220 nt, the G112C run-off exceeded the theoretical maximum possible run-off by approximately 40 nt. This G112C run-off artefact was similar in size to the 261 bp dsDNA template. It was suggested[3] the artefact could be the result of 32P-αATP labelling of template DNA as a minor activity of RNAP. A series of experiments were carried out to explore the hypothesis:

- The 3'end of the transcription template was shortened by 19 bp so that template run-off would occur at 162 nt. Transcripts generated from this template decreased in size proportionally to the WT run-off, remaining approximately 40 nt longer than theoretically possible (see Appendix Figure 10.X).
- Following elongation and heat denaturation, transcription reactions were subjected to DNase treatment to degrade any potentially labelled dsDNA template, but the G112C artefact remained unaffected (see Appendix Figure 10.XI).
- RNAP was incubated with WT template with 32P-αATP in the absence of any other rNTPs for 20 minutes to select for any potential template labelling activity, but no affect was observed (see Appendix Figure 10.XI).
- Both template vectors as well as PCR amplified DNA templates were sequenced and found correct.
- A transcription time-course was carried out with the G112C template over a period of 20 minutes to explore potential deviations in transcription pausing pattern (see Appendix Figure 10.XII). The pausing pattern appeared identical between the G112C and WT template up to the riboswitch intrinsic terminator. Besides the differences between the WT ~130 nt attenuation and discrete G112C attenuated transcripts (Figure 7.IV) there were no transcripts larger apart from discrepant the G112C artefact.

[3] Personal communication, Professor Nikolay Zenkin, Newcastle University (2017).

Collectively, investigations indicated the G112C artefact was occurring within the *rpfB* 5'UTR between the riboswitch intrinsic terminator and the template 3'terminus. It was therefore decided that inclusion of a terminal synthetic terminator could negate requirement of a template run-off by using synthetic termination as an indication of riboswitch read-though. Czyz *et al.* (2014) demonstrated the *synB* synthetic terminator to be a strong intrinsic terminator with both *E. coli* RNAP and recombinant *Mycobacterium bovis* BCG RNAP. Therefore the *synB* synthetic terminator was cloned into the template 3'terminus to generate a revised 302 bp template theoretically capable of riboswitch-mediated attenuation at ~130 nt, synthetic termination at 211 nt (according *synB* termination described by Czyz *et al.* or possible template run-off at 222 nt).

It was thought the presence of a terminal synthetic terminator would limit the maximum transcript length of transcript, but the G112C artefact remained unaffected and produced a *synB*-terminated transcript approximately 40 nt longer than theoretically possible (see Figure 7.IV, right). It is probable the cause of the G112C artefact occurs between the site of riboswitch-mediated attenuation and synthetic termination. The process by which the G112C artefact arises remains unknown (see section 7.4). Regardless of the G112C artefact, it is evident the riboswitch conformational mutants corroborated the results observed *in vivo* (see section 5.5). Both the anti-terminator stabilising U6C and poly(U) tract disrupting U117-119A substitutions resulted in discernible decreases in riboswitch-mediated attenuation and increased read-through. Together these results strongly indicate the riboswitch intrinsic terminator is functional and sufficient to actuate riboswitch-mediated transcription attenuation.

7.2.5 Transcriptional road-blocking validates intrinsic terminator function

A transcription road-blocking assay was devised to ascertain whether the riboswitch attenuated transcription through intrinsic termination or long-lived transcriptional pausing. The transcription template was biotinylated on the 5'end of the antisense strand by PCR amplification with a 5'biotinylated reverse primer (see section 2.9.2). Purified biotinylated template was coupled to streptavidin coated magnetic beads (Invitrogen) to create a biotin-streptavidin-bead conjugate. This conjugate presents a mechanical road-block to TECs, preventing template run-off so that TECs to remain bound to the template without

RNAP dissociating. However, TECs subject to intrinsic termination would dissociate from the template releasing terminated transcript. Template-bound TECs can be separated from reaction supernatant using a strong magnet, allowing terminated transcripts to be isolated.

The 261 bp biotinylated template was theoretically capable of riboswitch-mediated attenuation at ~130 nt, read-through to the biotin-streptavidin road-block at ~166 nt assuming a 30 nt RNAP footprint with a 7-9 nt DNA:RNA hybrid upstream from the centre of the footprint (Darst 2001; Greive & von Hippel 2005; Frieda & Block 2012) or possible template run-off at 181 nt if bead and template uncoupled during the reaction (see Figure 7.V). Reactions were carried out in the presence and absence of NusA which is known enhance attenuation at intrinsic terminators (Ha *et al.* 2010). Following a 5 minute elongation period, reaction supernatant and template-bound TECs were separated with a strong magnet and isolated. Aliquots of un-separated reactions were also taken as a control.

Although NusA did not appear to enhance intrinsic termination, Figure 7.V demonstrates ~130 nt attenuated transcript was detected in both supernatant samples (sup), indicating riboswitch-mediated intrinsic termination as released transcript. Conversely, pelleted samples (pel) indicated some TECs remained bound to the bead-coupled template collected by the magnet, suggesting a long-lived pause at the riboswitch. The road-blocked ~166 nt transcript confirms TECs remained template-bound. However, the presence of a larger transcript at ~180 nt indicated probable template run-off, possibly due to template uncoupling during the reaction as it was detected in both the supernatant and un-separated (mix) samples. At the bottom of the gel, halted TEC abortive transcripts were present in the supernatant samples only, confirming these released transcripts were abortive products.

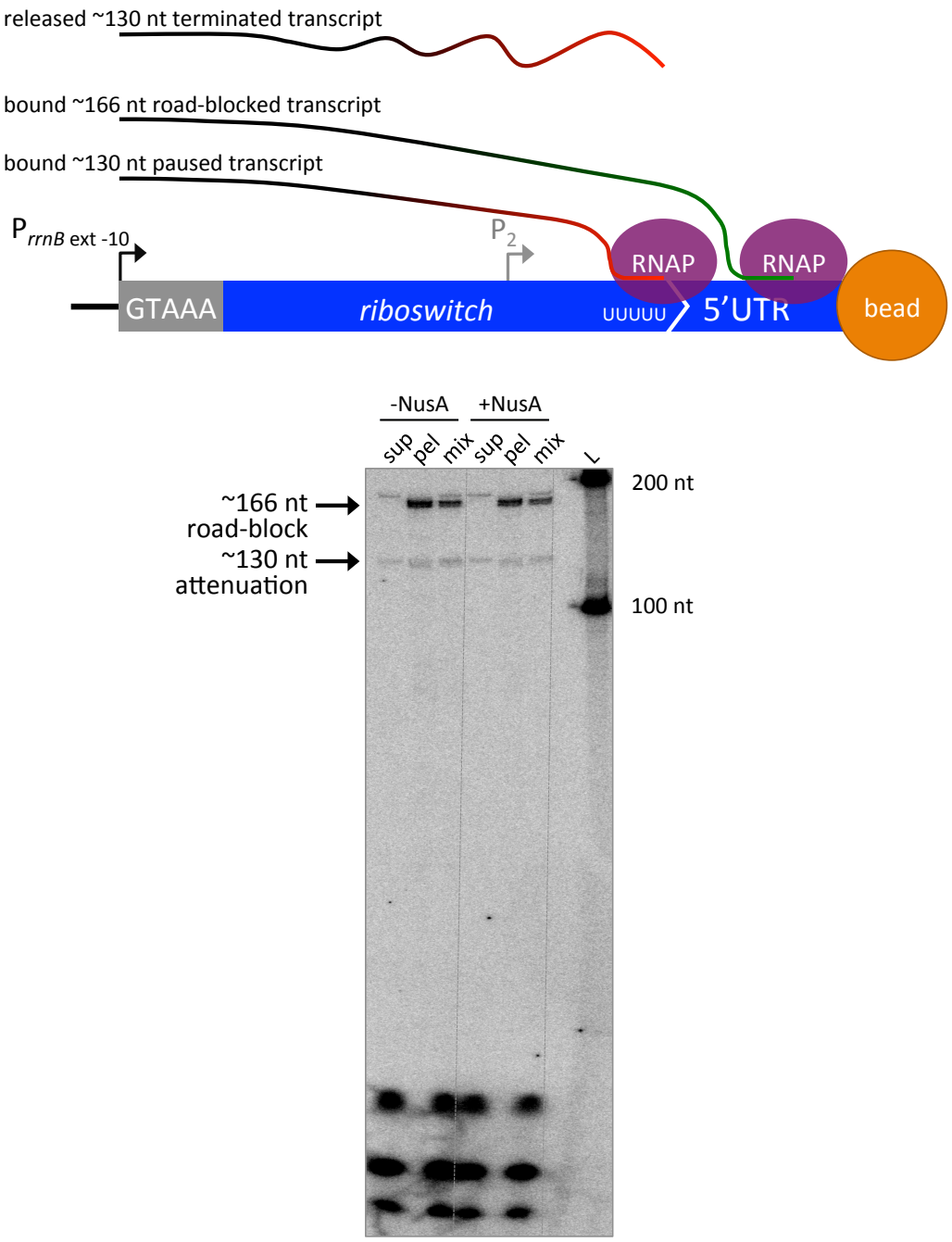


Figure 7.V *E. coli in vitro* transcription road-blocking
Biotin-streptavidin road-blocking of single-round *in vitro* transcription of the riboswitch. Left) reactions carried out in the absence of NusA; supernatant (sup), pelleted magnet-bound template (pel), un-separated reaction mixture (mix). Right) reactions carried out in the presence of NusA. Reactions were carried out at 30°C to decrease processivity, with 17 nM *E. coli* RNAP (NEB), initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Reactions were carried out ± 85 nM NusA (NEB). Elongation was resumed with 50 µM rNTPs, incubating for 5 min until separating reactions with strong magnet to harvest supernatant and pellet samples before stopping reactions with the addition of formamide loading dye. Road-blocked template was amplified with oligonucleotides 5.44 and 8.43 (see Appendix section 10.3).

Collectively the road-block assay results demonstrated the riboswitch attenuates transcription through both intrinsic terminator and long-lived pausing. However, the riboswitch poly(U) tract is imperfect, containing an interrupted series of uracil residues. Whilst this may be sufficient for intrinsic termination with mycobacterial RNAP, it may not be as effective in terminating transcription with *E. coli* RNAP.

7.3 Preliminary application of *M. tuberculosis* RNAP

By using commercially manufactured *E. coli* enzyme as a consistent and highly purified source of RNAP, it was possible to develop and optimise the *in vitro* transcription system described above. However, as the riboswitch is encoded by mycobacteria it was desirable to use a cognate *in vitro* transcription system. Recombinant *M. tuberculosis* RNAP holoenzyme was generously provided by Professor Sivaramesh Wigneshweraraj (Imperial College London). This enabled the preliminary application of *M. tuberculosis* to the *in vitro* transcription system.

The recombinant *M. tuberculosis* RNAP was provided as a purified enzyme generated by co-expression in *E. coli*. The holoenzyme was essentially produced as described by Banerjee *et al.* (2014) consisting of *M. tuberculosis* sigma factor (SigA) in complex with RNAP core ($\alpha_2\beta\beta'$) but lacking the *rpoZ* or omega (ω) subunit. Whilst this enzyme has been shown to be transcriptionally functional (Banerjee *et al.* 2014, 2015; Sharma *et al.* 2015; du Plessis *et al.* 2017), complex stability and function has been disputed in recent studies (see section 7.4). Nevertheless, *M. tuberculosis* RNAP subunits have been shown to have roughly only 50% conservation to *E. coli* subunits (du Plessis *et al.* 2017). Therefore it was decided it would be beneficial to experiment with transcription of the riboswitch using recombinant *M. tuberculosis* RNAP, albeit lacking the ω subunit. The *in vitro* transcription system was optimised for *M. tuberculosis* RNAP, identifying suitable nucleotide concentrations and buffering conditions (see section 2.10).

7.3.1 *M. tuberculosis* RNAP recognises riboswitch conformational mutants

To complement experiments carried out using *E. coli* RNAP (see Figure 7.IV) the conformational equilibrium of the riboswitch was explored using *M. tuberculosis* RNAP. The same mutant templates containing the terminator conformation stabilising G112C, anti-terminator conformation stabilising U6C and terminator poly(U) tract disrupting U117-119A substitutions were used to generate transcripts in single-round *in vitro* transcription reactions.

Figure 7.VI demonstrates *M. tuberculosis* RNAP was capable of riboswitch-mediated termination, generating the same ~130 nt transcript as observed using *E. coli* RNAP. In comparison to *E. coli* RNAP (see Figure 7.IV), the *M. tuberculosis* RNAP appeared to recognise the riboswitch intrinsic terminator more efficiently. The terminator conformation stabilising G112C substitution exhibited near complete termination whilst the intrinsic terminator poly(U) tract disrupting U117-119A substitutions resulted in near complete read-through. This strengthened the hypothesis that the *E. coli* RNAP is comparatively less efficient at intrinsic termination at the riboswitch intrinsic terminator due to the imperfect poly(U) tract. Unfortunately *M. tuberculosis* RNAP also produced the same G112C artefact, generating a run-off transcript longer than theoretically possible (grey arrow, Figure 7.VI). Regardless, this data further substantiates the conclusion that the riboswitch mediates transcription through intrinsic termination to regulate downstream gene expression.

Interestingly, the reproducibility of the ~130 nt riboswitch attenuated transcript and template run-off between the *M. tuberculosis* and *E. coli* holoenzymes indicated transcription was singularly promoted at the P_{rrnB} TSS in both systems. Whilst *E. coli* promoter recognition requires both -10 and -35 elements, *M. tuberculosis* promoters often have poor -35 element identity (Cortes *et al.* 2013). As the *M. tuberculosis* holoenzyme contains SigA for promoter recognition, a lack of any P_2 promoted transcripts suggests the P_2 promoter likely requires additional transcription factors to allow initiation. Whilst it is possible extended -10 P_{rrnB} out-competes P_2, this finding supports previous observations of the P_2 promoter *in vivo* where P_2 was found to have a much lower activity during exponential growth phases compared to P_1 (see section 5.2).

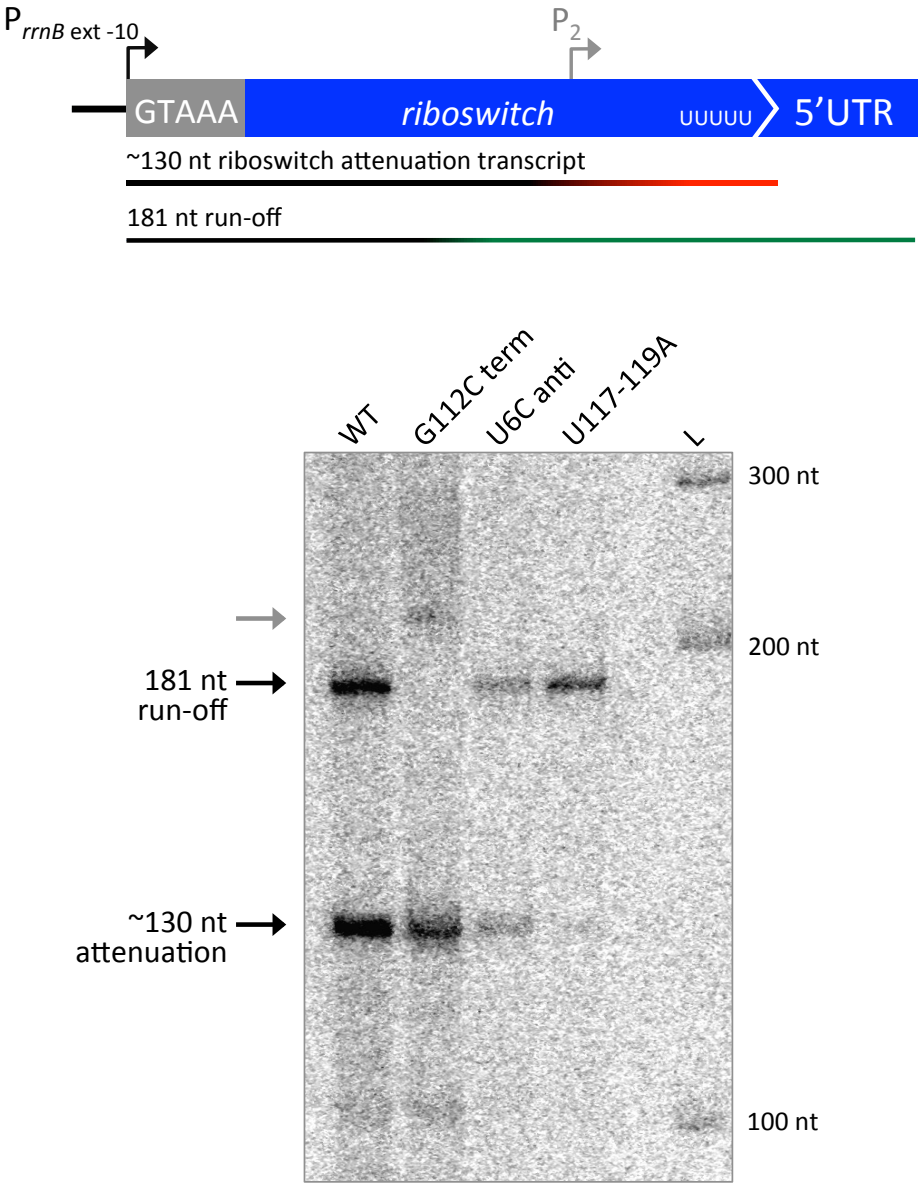


Figure 7.VI *M. tuberculosis in vitro* transcription of riboswitch mutants
Single-round *in vitro* transcription of riboswitch conformational mutants. From left to right; WT riboswitch template, followed by the G112C terminator stabilising substitution, the U6C anti-terminator stabilising substitution and U117-119A intrinsic terminator disrupting substitution. Grey arrow denotes G112C artefact. Reactions were carried out at 37°C with 400 nM *M. tuberculosis* RNAP, initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Elongation was resumed with 200 µM rNTPs, incubating for 20 min before stopping reactions with the addition of formamide loading dye. Run-off template amplified with oligonucleotides 5.31 and 5.23 (see Appendix section 10.3).

7.3.2 *M. tuberculosis* exhibits a different transcriptional pausing pattern

Differences in polymerase processivity have been found to influence transcriptional pausing and folding of nascent RNA, therefore it may be important to match riboswitches with suitable polymerase for heterologous *in vitro* transcription systems (Steinert *et al.* 2017). Distantly related polymerases may generate or interact with intermediate nascent RNA structures differently (Zhang & Landick 2016). This could interfere with regulatory function as aptamer folding may be co-ordinated with ligand recognition or pausing may be required for intrinsic termination. The riboswitch was transcribed *in vitro* in a batch single-round reaction using *M. tuberculosis* RNAP, taking aliquots periodically over a 20 minute elongation period and mixed with formamide loading dye to inhibit further transcription.

Figure 7.VII shows *M. tuberculosis* RNAP rapidly (<10 seconds) transcribed to the end of the template, as indicated by the abundance of synthetically terminated transcript at ~211 nt. By comparison, this is much faster than the *E. coli* time-course (see Figure 7.III), however the *E. coli* RNAP was intentionally constrained to slow elongation rate, limiting available rNTP concentration and decreasing reaction temperature to mimic the slower processivity of the *M. tuberculosis* RNAP. Even in the absence of *M. tuberculosis* NusA, the *M. tuberculosis* RNAP *in vitro* transcription system appears to be a more relevant system for riboswitch transcription. On comparison, the 41 and 43 nt dominant pauses observed using *E. coli* RNAP were completely absent in the *M. tuberculosis* RNAP reaction (grey arrows, Figure 7.VII). Rather a distinct long-lived pause was seen at 36 nt (black arrow, Figure 7.VII). This discrepancy in pausing pattern indicated the riboswitch interacts differently with *E. coli* and *M. tuberculosis* polymerases.

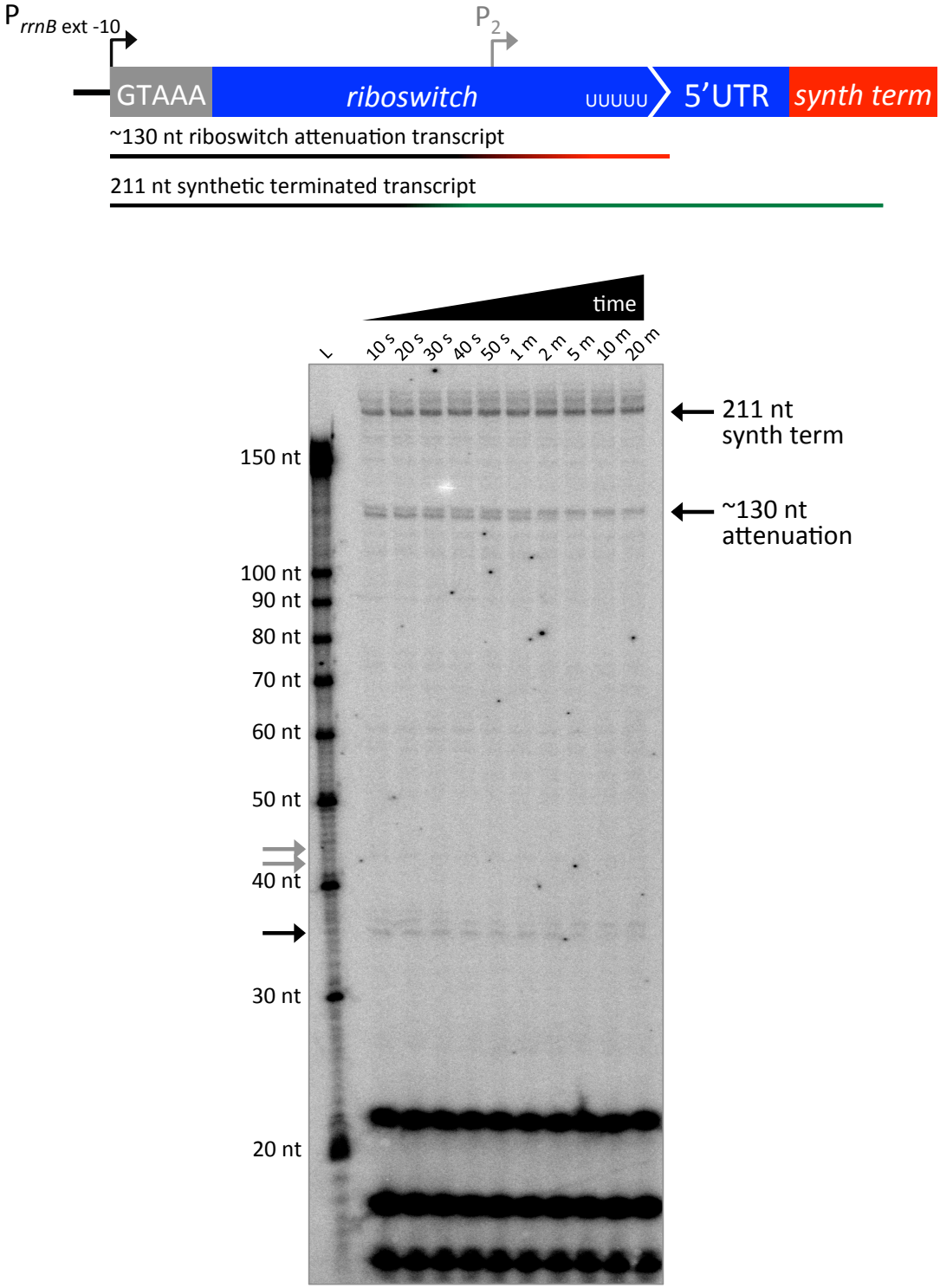


Figure 7.VII *M. tuberculosis in vitro* transcription time-course
Single-round *in vitro* transcription time-course of the riboswitch. Grey arrows denote the 41 and 43 nt dominant pauses observed with *E. coli* RNAP in Figure 7.III. Black arrow denotes *M. tuberculosis* specific pausing. Reactions were carried out at 37°C, with 400 nM *M. tuberculosis* RNAP, initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Elongation was resumed with 200 µM NTPs, incubating for up to 20 min before stopping reactions with the addition of formamide loading dye. Synthetic terminator template amplified with oligonucleotides 5.44 and 8.43 (see Appendix section 10.3).

7.4 Chapter discussion

Using commercially produced *E. coli* holoenzyme, it was possible to develop and optimise an *in vitro* transcription system capable of investigating the *rpfB* riboswitch regulatory control mechanism at the transcriptional level. A template was devised, capable of pausing transcription following initiation to allow the formation of halted TECs (Figure 7.II). This template demonstrated robust reproducibility and efficient promoter specific initiation using the *M. smegmatis* ribosomal *rrnB* operon P_1 promoter with extended 'TGN' -10 element. By withholding UTP from reactions and initiating with GpU dinucleotide, TECs were uniformly halted at U+11 (U+6 relative to *rpfB* P_1 TSS). Upon addition of the full complement of rNTPs, halted TECs resumed elongation as a single-round transcription event.

Synchronicity of transcription was exploited to explore TEC processivity during riboswitch transcription over time. *E. coli* holoenzyme time-course series identified a reproducible transcriptional pausing pattern for the riboswitch (Figure 7.III). Dominant long-lived pauses were observed at 41 and 43 nt (G+36 and G+38 relative to *rpfB* P_1 TSS). Intermediate 41 and 43 nt transcript sequences were studied in structural models (data not shown), however no strong secondary structures were predicted (Zuker 2003), indicating these pauses are likely caused by interaction with the *E. coli* RNAP itself. Indeed, when a similar time-course series was carried out using *M. tuberculosis* RNAP these 41 and 43 nt pauses did not occur, rather a separate long-lived pause was observed at 36 nt (Figure 7.VII).

This discrepancy in pausing pattern highlights the importance of using a cognate transcription system for riboswitch investigation. *E. coli* RNAP subunits have been shown to have roughly only 50% conservation with *M. tuberculosis* subunits (du Plessis *et al.* 2017) which exhibits a lower processivity than *E. coli* (Herrera-Asmat *et al.* 2017). Template and transcript nucleotide composition can also influence pause duration (Zhang & Landick 2016). This is an important consideration when using polymerase from *E. coli*, a relatively A:T rich organism, to transcribe a G:C rich riboswitch. Not only can varied processivity influence transcriptional pausing and folding of nascent RNA, but distantly related polymerases can interact differently with nascent RNA structures (Zhang & Landick 2016; Steinert *et al.* 2017).

In this study, rNTP concentration and reaction temperature were purposely constrained to simulate reduced processivity with *E. coli* RNAP. This was somewhat successful, producing a series of discrete transcriptional pauses that were further enhanced with the addition of NusA (Figure 7.III), a transcription factor known to increase duration of pausing (Artsimovitch & Landick 2000). However, as *M. tuberculosis* RNAP produced a different pausing pattern, the relevance of these *E. coli* RNAP pauses is contentious, especially if they represent different intermediate nascent RNA structures. For riboswitches especially, folding intermediates can be crucial for ligand interaction (Nechooshtan *et al.* 2009; Frieda & Block 2012; Perdrizet *et al.* 2012).

Nevertheless, with application of NusA, *E. coli* RNAP demonstrated a decreased transcription rate. Co-transcriptional ligand interaction can be important for transcriptional riboswitches and influenced by elongation rate (Wickiser *et al.* 2005). A 1:5 molar ratio of *E. coli* NusA was added to the *M. tuberculosis* RNAP time-course in an attempt to replicate this effect (data not shown). Whilst the RNAP interacting N-terminal of NusA is well conserved between the two species, the C-terminal is quite different (Arnvig *et al.* 2004). Consequently, *E. coli* NusA did not decrease *M. tuberculosis* RNAP processivity. In future it would be beneficial to purify *M. tuberculosis* NusA, which would be more biologically relevant to *M. tuberculosis* RNAP. As a homologous *in vitro* transcription system this would likely facilitate nascent RNA folding to affect pausing and termination (Beuth *et al.* 2005).

Despite the discrepant pausing patterns observed, the riboswitch conformational mutants demonstrated both enzymes were capable of recognising the riboswitch sequence to generate similar transcripts. Comparing the *E. coli* RNAP results (Figure 7.IV) with the *M. tuberculosis* (Figure 7.VI), both polymerases recognised the riboswitch intrinsic terminator to yield identical ~130 nt attenuated transcripts. In both systems, the terminator conformation stabilising G112C substitution (C+117 relative to the P_{rrnB} TSS) resulted in increased attenuation and the intrinsic terminator poly(U) tract disrupting U117-119A substitution (U+122 to U+124 relative to the P_{rrnB} TSS) resulted in increased read-through.

Interestingly these trends were even more apparent in the *M. tuberculosis* system where the G112C substitution resulted in near total attenuation and the U117-119A substitution in near total read-though. This further highlights the need for a cognate *in vitro* transcription

system. It is likely *E. coli* RNAP did not efficiently attenuate transcription at the intrinsic terminator due to the imperfect poly(U) tract. Whereas *M. tuberculosis* RNAP is more promiscuous and may terminate at imperfect poly(U) tracts of ≥4U residues, even if subsequent nucleotides contain other bases (Czyz *et al.* 2014), as in the *rpfB* riboswitch (UUUUGCUUUUGUU).

The transcriptional road-block assay (Figure 7.V) established the proposed riboswitch intrinsic terminator is functional and sufficient to actuate riboswitch-mediated transcription attenuation. Riboswitch attenuated transcript was detected in both the supernatant and template-bound fractions of the reaction, indicating attenuation occurs by both intrinsic termination and long-lived transcriptional pausing at the terminator. Given the inherent instability that weak A:U basepairs of the DNA:RNA impose on the paused complex, such long-lived pause would likely result in further termination over time through RNAP dissociation.

Although *E. coli* RNAP demonstrated NusA had no effect on intrinsic termination efficiency (Figure 7.V), it possible an effect could be observed using *M. tuberculosis* RNAP and NusA. Alternatively, intrinsic termination might require additional transcription factors. NusG is known to stabilise the RNAP clamp, increasing processivity to enhance intrinsic termination through hyper-translocation of nascent RNA from RNAP active site (Czyz *et al.* 2014). RNAP can freely utilise both NusA and NusG co-operatively as they bind at different sites (Artsimovitch & Landick 2000). It may be that both of these factors are required for correct processivity to fold the nascent RNA for ligand interaction or termination.

Despite extensive investigation, no explanation was found for the G112C transcription artefact observed in both *E. coli* (Figure 7.IV), and *M. tuberculosis* systems to a lesser extent (Figure 7.VI). In all instances, templates with the G122C substitution generated a transcript approximately 40 nt longer than the theoretical possible maximum. Perhaps this was the result of *in vitro* transcription in the absence of additional transcription factors. Time-course experiments showed that the WT and G122C templates exhibited identical transcriptional pausing patterns (Appendix Figure 10.XII), indicating the G112C artefact occurs downstream of the riboswitch intrinsic terminator sequence. Potentially, non-template nucleotide incorporation could have occurred by backtracking following pausing.

Long-lived transcriptional pausing can induce backtracked transcription, a mechanism for rescuing stalled transcription complexes whereby the nascent RNA is translocated out the elongation complex through a polymerase secondary exit channel (Marr & Roberts 2000). Transcription factors GreA and GreB stimulate endonucleolytic cleavage of the extruded RNA and transcription is subsequently resumed with the nascent RNA 3'terminus within the RNAP active site (Tetone *et al.* 2017). Gre factor homologues have been identified in mycobacteria (China *et al.* 2012). In the absence of Gre factor, backtracking without nascent RNA cleavage could have resulted in non-template nucleotide addition, lengthening the G112C transcript. However, adding additional factors to the transcription reaction would only serve to increase system complexity for investigation of an experimental artefact that likely bears little relevance *in vivo*, as demonstrated by the G112C reporter construct in previous chapters (see section 5.5).

With evidence indicating *M. tuberculosis* RNAP transcribes and interacts with the riboswitch uniquely, it was decided the *M. tuberculosis in vitro* transcription system would be the more relevant system to use for candidate ligand screening. Preliminary experiments were conducted using the *M. tuberculosis* system with the candidate ligand SAM at a range of concentrations but no change in riboswitch-mediated attenuation was observed (data not shown). Certainly this may indicate SAM is not recognised as ligand by the riboswitch, despite rational reasoning. However, it may also accentuate inadequacies with the *M. tuberculosis* system itself.

Recently Herrera-Asmat *et al.* (2017) demonstrated omission of the ω subunit from *M. tuberculosis* core polymerase results in complex degradation during protein purification, incorrect subunit stoichiometry and β' subunit shearing (Herrera-Asmat *et al.* 2017). Whilst complexes retained DNA binding capacity, transcriptional activity was limited. Therefore, in the absence of the ω subunit, molar concentration of transcriptionally functional RNAP cannot be accurately determined. In the present study, single-round transcription reactions with *M. tuberculosis* RNAP lacking the ω subunit exhibited varied processivity in preliminary experiments whereby TECs took between 10 seconds to 40 minutes to transcribe the entire template (data not shown). This is indicative of a heterologous polymerase preparation, likely due to subunit degradation during protein purification (Herrera-Asmat *et al.* 2017).

Although there is a precedence in literature for excluding ω during recombinant expression (Banerjee *et al.* 2014), subsequent studies have chosen to include it in core polymerase expression (Herrera-Asmat *et al.* 2017; Lin *et al.* 2017). Indeed there is increasing evidence of its importance as part of the polymerase complex as deletion of *rpoZ* leads to β' misfolding and degradation in *M. smegmatis* (Mathew *et al.* 2005) and *E. coli* (Ghosh, Ishihama & Chatterji 2001). In *rpoZ* deletion strains of *Staphylococcus aureus* not only did RNAP lacking the ω subunit exhibit decreased sigma factor recruitment specificity, but *rpoZ* deletion was found to affect biofilm formation, stress adaptation and the stringent response (Weiss *et al.* 2017).

During transcription, nascent RNA experiences specific interactions with exit channel and surface of RNAP which may in turn undergo transcriptional pauses necessary for correct folding (Lutz *et al.* 2014; Zhang & Landick 2016). In the present study *M. tuberculosis* RNAP and riboswitch nascent RNA appeared to interact uniquely, however results might be quite different using *M. tuberculosis* RNAP inclusive of the ω subunit. For this reason, no further candidate ligands were tested, as results of such assays would have been inconclusive.

For further experimentation in the future it would be essential to purify protein for the entire canonical core polymerase ($\alpha_2\beta\beta'\omega$) as well as transcription factors, NusA in particular. Following optimisation of such a system, candidate ligands could be explored, principally SAM precursors/derivatives of the methyl cycle (Parveen & Cornell 2011), as well as other frequently identified riboswitch ligands such as nucleotide derivatives (Nelson *et al.* 2013; Kellenberger *et al.* 2015; St-Onge *et al.* 2015) and cellular metabolites (Frieda & Block 2012; Polaski *et al.* 2017). However, as a regulatory 'on-switch', ligand identification for the *rpfB* riboswitch may be more challenging, especially if the ligand is a secondary messenger indirectly associated with downstream gene expression, which has proven to be the case for some orphaned riboswitches (Meyer *et al.* 2011).

7.5 Chapter conclusion

In conclusion, a robust single-round *in vitro* transcription assay has been developed and optimised to evaluate the riboswitch mechanisms of regulatory control on a transcriptional level. Conformational mutants have demonstrated yet further the capacity for the riboswitch to affect transcription attenuation, likely through conformational switching. Transcriptional road-blocking established attenuation is achieved through intrinsic termination and long-live transcriptional pausing at the riboswitch intrinsic terminator. Time-course series demonstrated transcription of the riboswitch undergoes discrete transcriptional pausing. Moreover, results indicated *M. tuberculosis* RNAP interacts uniquely with the riboswitch, which may be crucial in cognate ligand identification in the future.

8 Discussion and conclusions

8.1 Overall discussion

Mycobacterium tuberculosis can reactivate to develop active disease years after the primary infection, which is thought to be achieved through resuscitation from dormancy by resuscitation promoting factors (Rosser *et al.* 2017). *M. tuberculosis* encodes five resuscitation promoting factors *rpf*(*A-E*). At present, only two riboswitches have been described in *M. tuberculosis*; the cobalamin sensing *metE* methionine synthase riboswitch (Warner *et al.* 2007) and the c-di-AMP sensing[4] *ydaO* riboswitch homologue upstream of *rpfA* (Block, Hammond & Breaker 2010; Arnvig & Young 2012; Nelson *et al.* 2013; St-Onge *et al.* 2015). The five *M. tuberculosis* Rpfs are thought to be functionally redundant to a certain extent (Zhonghe & Zhang 1999; Tufariello, Jacobs & Chan 2004; Downing *et al.* 2005; Tufariello *et al.* 2006), yet it is intriguing that two of the five *rpfs* genes have now been found to be subject to riboswitch-mediated regulation by very different riboswitches. The work described in this thesis is the first instance of a riboswitch having been identified upstream of *rpfB* in any species.

[4] Personal communication, Dr Galina V. Mukamolova, University of Leicester (2017).

In previous chapters a novel transcriptional riboswitch was characterised upstream of an *rpfB-ksgA* bi-cistron. This riboswitch appears be conserved in pathogenic species alone, drawing a potential link to pathogenesis. Riboswitch-mediated attenuation was observed through northern blot analysis, mapped by 3'rapid amplification of cDNA ends (RACE) and manipulated both *in vitro* and *in vivo*. Transcriptional road-blocking was used to demonstrate this attenuation is achieved through intrinsic termination at the riboswitch terminator. Moreover, *in vitro* transcription time-course series revealed the riboswitch interacts uniquely with *M. tuberculosis* RNA polymerase (RNAP) to generate a series of discrete pauses that may be important for ligand binding or nascent RNA folding.

Structural modelling of the *rpfB* riboswitch predicts two possible conformations: a terminator conformation and an anti-terminator conformation (Zuker 2003). The stable folded free energy of these conformations predicts the riboswitch exists in the terminator conformation by default, as a transcriptional 'on-switch', where it is presumed ligand interaction would enhance transcriptional read-through. Close inspection of the two conformations shows they operate under a paradigm of riboswitch function as a mutually exclusive switch (Batey 2011; Millman *et al.* 2017). Formation of the intrinsic terminator stem-loop precludes the formation of the anti-terminator conformation and *vice versa* (see Figure 8.I).

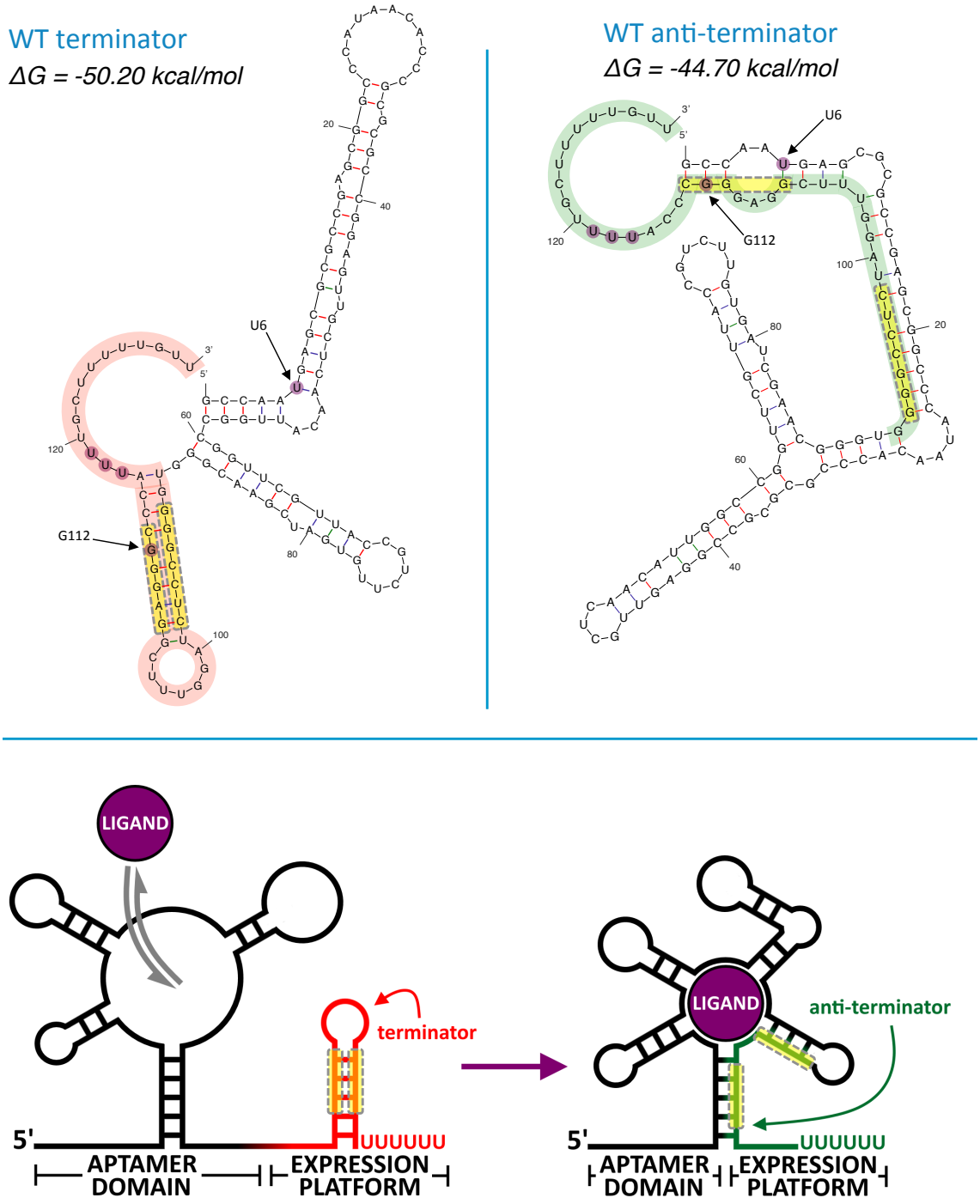


Figure 8.I a mutually exclusive switching mechanism
The top diagrams show the predicted *rpfB* riboswitch conformations with the stem region of the intrinsic terminator stem-loop highlighted in yellow (Zuker 2003). The bottom diagram demonstrates how one conformation precludes the formation of another in a riboswitch mutually exclusive switch (adapted from Batey 2011).

Reverse transcription PCR (RT-PCR) demonstrated co-transcription of *rpfB* and *ksgA* in *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG (see section 4.2); therefore pathogens that encode the riboswitch also subject this bi-cistron to its regulatory control. The function of the *rpfB-ksgA* bi-cistron is uncertain; perhaps it is a regulatory link that enables a co-ordinated restoration of active growth, achieved through resuscitation by the RpfB cell wall hydrolase and ribosome biogenesis by the KsgA methyltransferase. Whilst it has previously been observed that actinobacteria frequently encode a cell wall enzyme proximal to *ksgA* (Ravagnani, Finan & Young 2005), it remains unclear as to why pathogenic mycobacteria would also encode riboswitch-mediated regulation of this arrangement. The *rpfB* riboswitch cognate ligand may be particularly relevant to pathogenesis. Perhaps it is a host-derived signal that indicates an opportunistic environment in which resuscitation from dormancy would be beneficial to the pathogen. In *Listeria monocytogenes* the *prfA* 5'UTR thermosensor allows expression of virulence genes under such conditions, revealing the SD sequence by melting at 37°C inside the infected host (Johansson *et al.* 2002). If the *rpfB* riboswitch indicates host presence to mycobacteria, then it would crucial for future research to focus on riboswitch expression within infection models such as macrophage tissue culture or animal models.

Northern blot analysis of *M. tuberculosis* exponential growth phase total RNA indicated roughly an equal ratio of riboswitch attenuation and read-through *in vivo* (see section 5.2). As a predicted transcriptional 'on-switch' this would suggest the ligand might be present during culture to a certain extent. Perhaps the ligand is therefore a secondary signalling molecule that is up-regulated by stimuli indicative of environments conducive to resuscitation. Certainly this could account for some of the challenges that have been encountered in attempting to identify the cognate ligand. Indeed it has been theorised that several orphan riboswitches likely interact with secondary signalling molecules in the same manner as c-di-AMP and c-di-GMP sensing riboswitches (Sudarsan *et al.* 2008; McCown *et al.* 2017).

Dormant *M. tuberculosis* exhibit very thick heavily cross-linked cell walls that likely restrict the cells from normal division and growth. It is thought that action of resuscitation promoting factors is required for digestion and restoration of normal cell wall homeostasis

(Hett & Rubin 2008). The RpfB cell wall hydrolase interacts with another hydrolase, RipA, for increased digestion of the cell wall (Hett *et al.* 2007). It has been shown that RpfB/RipA cell wall digests can act as an extremely potent resuscitation signal for other dormant mycobacteria (Nikitushkin *et al.* 2015). Perhaps this is achieved through the generation of secondary signalling molecules as digest fragments bind at the cell surface to the PknB serine-threonine kinase (Nikitushkin, Demina & Kaprelyants 2016). Potentially such secondary signalling molecules could be identified by the *rpfB* riboswitch as a signal of neighbouring cells undergoing resuscitation to subsequently express *rpfB-ksgA*.

An alternative theory that has been explored in this thesis is that the riboswitch may be more applicable to *ksgA* expression rather than *rpfB*. As a methyltransferase KsgA utilises S-adenosyl-L-methionine (SAM) as an essential methyl donor for the methylation of ribosomal rRNA (Connolly, Rife & Culver 2008). Not only do mycobacteria encode an increased number of SAM synthases and methyltransferases (Cole *et al.* 1998; Berger & Knodel 2003), but they also use the molecule for biosynthesis of cell envelope mycolic acids and adhesion modifications (Parveen & Cornell 2011). Certainly it is feasible that SAM would serve as an excellent indicator of advantageous growth conditions for the regulation of resuscitation. Disappointingly, *in vitro* transcription of the riboswitch in the presence of SAM did not show any changes in riboswitch-mediated attenuation that would have been indicative of ligand recognition (see section 7.2.2). This may be because the riboswitch is incapable of interacting with SAM, or perhaps in the absence of the ω subunit the *M. tuberculosis* RNAP does not interact with nascent RNA in a manner that would allow riboswitch interaction with SAM. This remains an area for future study.

Unlike the *rpfB* P_1 promoter, the P_2 promoter is encoded by both pathogens and non-pathogens. P_2 promoted transcripts are not subjected to the regulatory control of the riboswitch that is encoded upstream. In this study, promoter fusions demonstrated P_2 makes an unsubstantial contribution to *rpfB* expression during exponential growth in *M. tuberculosis* (see section 5.3). In addition, real-time quantitative PCR (RT-qPCR) of cDNA generated from biofilm extracts showed P_2 expression might well be increased in mature biofilms (see section 6.5). Not only does it seem that expression from the P_2 promoter likely requires the involvement of transcription factors or activators, but it appears that there are some conditions under which mycobacteria express the *rpfB-ksgA* bi-cistron independently

of riboswitch-mediated regulation. Indeed, non-pathogenic species do so in all conditions as they encode neither the riboswitch nor the P_1 promoter.

The regulatory control the riboswitch enables is specific to a subset of pathogenic mycobacteria, specifically the *M. tuberculosis* complex. Such tight regulation of the *rpfB* locus is further controlled by the expression of a potential asRNA. Although it was not possible to detect the asRNA by northern blot or map it with 3'RACE, reporter fusions indicated the P_{as} promoter to be transcriptionally active. Furthermore, RT-qPCR of biofilm cDNA demonstrated it has a relatively low expression (see section 6.5). Perhaps the asRNA functions to down-regulate spurious riboswitch read-through, either through complementary basepairing as a regulatory asRNA or to reduce transcription through polymerase clashing. An alternative theory that has been proposed in this thesis is that it may serve to occlude the *rpfB* ribosome binding site (RBS) that has been verified through translational reporters. RBS occlusion would present an additional translational checkpoint for *rpfB* expression. Maybe under advantageous conditions the P_{as} promoter is repressed for optimal riboswitch-mediated expression, but activated under disadvantageous conditions to rapidly down-regulate *rpfB* translation through RBS occlusion. Re-annotation of the *rpfB* translation start codon from ATG to the upstream TTG in *M. tuberculosis,* as identified in this study (see section 3.2), allows for a more typical view of the *rpfB* 5'UTR. The *rpfB* CDS 5'end is within a feasible distance from the verified RBS and makes the asRNA 5'end well placed for regulation of the region as described.

The riboswitch controls the expression of two classical drug targets involved in cell wall homeostasis and ribosome biogenesis. Riboswitches make excellent drug targets as they have evolved to bind a single cognate ligand with a high affinity and specificity using the limited repertoire of RNA nucleotides for secondary structure (Blount & Breaker 2006). As such their capacity to mutate and develop antibiotic resistance without affecting cognate ligand recognition is limited. The *rpfB* riboswitch therefore represents a highly selective drug target, specific to pathogenic mycobacteria. Such a target could be used for rational drug design, a development process largely absent for antimycobacterials currently (Kana *et al.* 2014).

Treatments capable of repressing resuscitation from dormancy could be capable of preventing the development of an active infection from latency. Indeed, an inhibitor of the RpfB enzyme was recently described that is capable of significantly decreasing the ability of *M. tuberculosis* to resuscitate (Demina *et al.* 2017). The riboswitch described in this thesis regulates the expression of both *rpfB* and *ksgA* at the transcriptional level. As a transcriptional 'on-switch' it is presumed that ligand interaction would enhance read-through into downstream genes. Given the highly specific nature of riboswitches, rational design of an inhibitor or competitor could therefore enable strong expression inhibition.

In the absence of empirically validated riboswitch structures, the riboswitch may adopt a rather different structure in the presence of bound ligand. It could therefore be contested that the riboswitch might function as an 'off-switch', terminating in the presence of ligand. Although this seems improbable as termination was observed during *in vitro* transcription reactions where it is unlikely any of the components present would serve as a ligand. Nevertheless, as a transcriptional 'off-switch' the *rpfB* riboswitch could still represent a viable drug target. In such a scenario ligand competitors could facilitate over-expression of the *rpfB-ksgA* bi-cistron, potentially preventing establishment of dormant populations during infection.

Research within this study used basic techniques to explore individual aspects that contribute to the expression of the *rpfB* locus as demonstrated in Figure 8.II. Prior to this study, global transcription start site mapping (TSSM) had been used to annotate putative promoters for the region (Cortes *et al.* 2013). These promoter elements have now been verified through reporter fusions. Similarly, only the annotated ATG start codon of *rpfB* had previously been investigated (Sharma *et al.* 2015), but through translational start site mapping (TLSM) this study has demonstrated the upstream TTG start codon is the correct 5'end of the CDS with a verified RBS upstream. Furthermore the theorised link between *rpfB* and *ksgA* (Ravagnani, Finan & Young 2005) has been substantiated through RT-PCR and identified as a bi-cistron. All of these findings are very likely to be reciprocated in other related mycobacteria species where the *rpfB* locus is highly conserved.

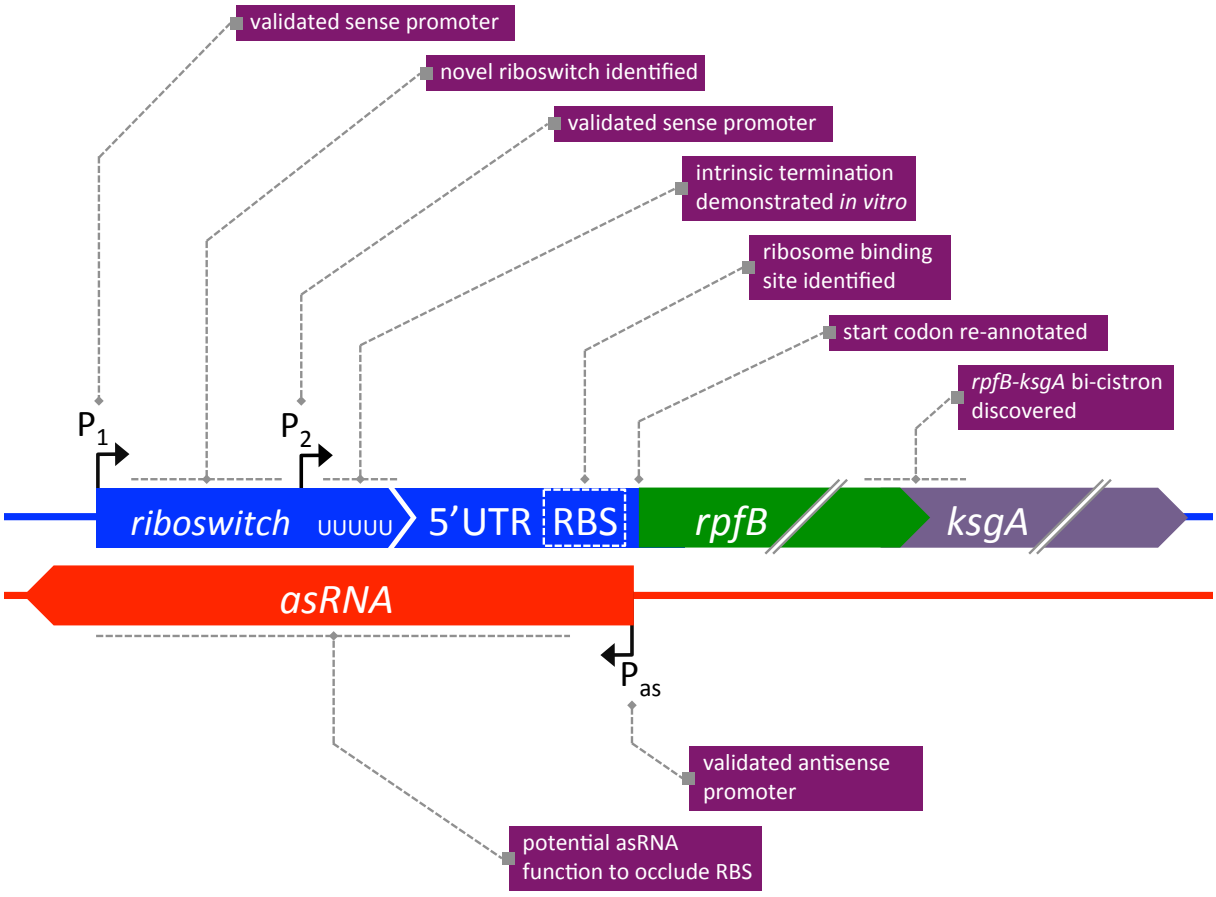


Figure 8.II diagrammatic summary of study findings

8.2 Future perspectives

Whilst this study has exposed many characteristics of the *rpfB* riboswitch that may qualify it as a riboswitch, the only assured test of riboswitch function is to demonstrate ligand dependent conformational switching. This can only be achieved with the discovery of the cognate ligand. In future, with the use of *M. tuberculosis* NusA and canonical RNAP inclusive of the ω subunit, the *in vitro* transcription system developed in this study could be used to screen candidate ligands capable of affecting transcription attenuation at the riboswitch intrinsic terminator.

Another approach that could be considered is the use of an *M. tuberculosis* transposon mutant library. With each conditionally non-essential gene knocked out through transposon insertion, transposon libraries transformed with a riboswitch reporter fusion could indicate increased or decreased riboswitch read-through through β-galactosidase expression on X-gal agar. Clones exhibiting differential expression could then be sequenced to identify the

genes contributing to altered expression. Potentially this could reveal genes or pathways associated with the ligand and allow for a more targeted application of the *in vitro* transcription assay screening.

Further focus on structural probing of the riboswitch would allow additional constraints to be applied to structural models to increase confidence in predicted structures. Structural probing could also be carried out on discovery of the cognate ligand to identify riboswitch regions crucial to ligand recognition. Such advancements would greatly aid the rational development of inhibitors or competitors.

By targeting different regions of the *rpfB* locus for RT-qPCR of the biofilm extract cDNA, it was apparent that the P_1 and P_2 promoters likely contribute to the expression of *rpfB* under different conditions. To further understand how *rpfB* is expressed it would be advantageous to conduct RT-qPCR on cDNA generated from multiple growth condition extracts. This would help elucidate conditions conducive to resuscitation or conditions where the riboswitch-mediated regulation is bypassed by P_2 promoted transcription.

8.3 Overall conclusion

Overall, this work has identified a novel riboswitch that regulates expression of an *rpfB-ksgA* bi-cistron at the transcriptional level through intrinsic termination. Only pathogenic mycobacteria encode this riboswitch. This riboswitch regulation of resuscitation from dormancy and ribosome biogenesis presents a strong candidate for the development of antimycobacterial therapeutics.

9 References

Agarwal N & Tyagi AK (2003) '*Role of 5'-TGN-3' motif in the interaction of mycobacterial RNA polymerase with a promoter of "extended -10" class*', *FEMS Microbiology Letters*, Vol. **225**, pp. 75–83.

Arnvig KB (2001) '*Transcription of rDNA in Mycobacteria*', *National Institute for Medical Research, Doctoral Thesis*.

Arnvig KB, Gopal B, Papavinasasundaram KG, Cox RA & Colston MJ (2005) '*The mechanism of upstream activation in the rrnB operon of Mycobacterium smegmatis is different from the Escherichia coli paradigm*', *Microbiology*, Vol. **151**, pp. 467–473.

Arnvig KB, Pennell S, Gopal B & Colston MJ (2004) '*A high-affinity interaction between NusA and the rrn nut site in Mycobacterium tuberculosis*', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. **101**, pp. 8325–8330.

Arnvig KB & Young DB (2009) '*Identification of small RNAs in Mycobacterium tuberculosis*', *Molecular Microbiology*, Vol. **73**, pp. 397–408.

Arnvig KB & Young DB (2012) '*Non-coding RNA and its potential role in Mycobacterium tuberculosis pathogenesis*', *RNA Biology*, Vol. **9**, pp. 427–436.

Artsimovitch I & Henkin TM (2009) '*In vitro approaches to analysis of transcription termination*', *Methods*, Vol. **47**, pp. 37–43.

Artsimovitch I & Landick R (2000) '*Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals*', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. **97**, pp. 7090–7095.

Artsimovitch I & Santangelo TJ (eds) (2015) '*Bacterial transcriptional control - methods and protocols*', Springer Protocols.

Banerjee R, Rudra P, Prajapati RK, Sengupta S & Mukhopadhyay J (2014) '*Optimization of recombinant Mycobacterium tuberculosis RNA polymerase expression and purification*', *Tuberculosis*, Vol. **94**, pp. 397–404.

Banerjee R, Rudra P, Saha A & Mukhopadhyay J (2015) '*Recombinant reporter assay using transcriptional machinery of Mycobacterium tuberculosis*', *Journal of Bacteriology*, Vol. **197**, pp. 646–653.

Batey RT (2011) '*Recognition of S-adenosylmethionine by riboswitches*', *Wiley Interdisciplinary Reviews: RNA*, Vol. **2**, pp. 299–311.

Bentrup KHZ & Russell DG (2001) '*Mycobacterial persistence: adaptation to a changing environment*', *Trends in Microbiology*, Vol. **9**, pp. 597–605.

Berger BJ & Knodel MH (2003) '*Characterisation of methionine adenosyltransferase from Mycobacterium smegmatis and M. tuberculosis*', *BMC Microbiology*, Vol. **3**, p. 12.

Beuth B, Pennell S, Arnvig KB, Martin SR & Taylor IA (2005) '*Structure of a Mycobacterium tuberculosis NusA–RNA complex*', *The EMBO Journal*, Vol. **24**, pp. 3576–3587.

Blattner FR, Plunkett GI, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Wayne Davis N, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B & Shao Y (1997) '*The complete genome sequence of Escherichia coli K-12*', *Science*, Vol. **277**, pp. 1453–1462.

Block KF, Hammond MC & Breaker RR (2010) '*Evidence for widespread gene control function*

by the ydaO riboswitch candidate', *Journal of Bacteriology*, Vol. **192**, pp. 3983–3989.

Blount KF & Breaker RR (2006) '*Riboswitches as antibacterial drug targets*', *Nature Biotechnology*, Vol. **24**, pp. 1558–1564.

Branda SS, Vik Å, Friedman L & Kolter R (2005) '*Biofilms: the matrix revisited*', *Trends in Microbiology*, Vol. **13**, pp. 20–26.

Brantl S (2007) '*Regulatory mechanisms employed by cis-encoded antisense RNAs*', *Current Opinion in Microbiology*, Vol. **10**, pp. 102–109.

Brantl S (2012) '*Acting antisense: plasmid- and chromosome-encoded sRNAs from Gram-positive bacteria*', *Future Microbiology*, Vol. **7**, pp. 853–871.

Brennan RG & Link TM (2007) '*Hfq structure, function and ligand binding*', *Current Opinion in Microbiology*, Vol. **10**, pp. 125–133.

Centers for Disease Control and Prevention & Kubica GP (1979) '*ID#: 5789*' Public Health Image Library.

Chauvier A, Picard-Jean F, Berger-Dancause J-C, Bastet L, Naghdi MR, Dubé A, Turcotte P, Perreault J & Lafontaine DA (2017) '*Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation*', *Nature Communications*, Vol. **8**, p. 13892.

China A, Mishra S, Tare P & Nagaraja V (2012) '*Inhibition of Mycobacterium tuberculosis RNA polymerase by binding of a Gre factor homolog to the secondary channel*', *Journal of Bacteriology*, Vol. **194**, pp. 1009–1017.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream M-A, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S & Barrell BG (1998) '*Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence*', *Nature*, Vol. **393**, pp. 537–544.

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D & Gagneux S (2013) '*Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans*', *Nature Genetics*, Vol. **45**, pp. 1176–1182.

Commichau FM, Dickmanns A, Gundlach J, Ficner R & Stülke J (2015) '*A jack of all trades: the multiple roles of the unique essential second messenger cyclic di-AMP*', *Molecular Microbiology*, Vol. **97**, pp. 189–204.

Connolly K, Rife JP & Culver G (2008) '*Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA*', *Molecular Microbiology*, Vol. **70**, pp. 1062–1075.

Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC & Dye C (2003) '*The growing burden of tuberculosis*', *Archives of Internal Medicine*, Vol. **163**, pp. 1009–1021.

Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R & Young DB (2013)

'*Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in Mycobacterium tuberculosis*', *Cell Reports*, Vol. **5**, pp. 1121–1131.

Czyz A, Mooney RA, Iaconi A & Landick R (2014) '*Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators*', *mBio*, Vol. **5**, pp. e00931-14.

Dambach M, Sandoval M, Updegrove TB, Anantharaman V, Aravind L, Waters LS & Storz G (2015) '*The ubiquitous yybP-ykoY riboswitch is a manganese-responsive regulatory element*', *Molecular Cell*, Vol. **57**, pp. 1099–1109.

Dambach MD & Winkler WC (2009) '*Expanding roles for metabolite-sensing regulatory RNAs*', *Current Opinion in Microbiology*, Vol. **12**, pp. 161–169.

Darst SA (2001) '*Bacterial RNA polymerase*', *Current Opinion in Structural Biology*, Vol. **11**, pp. 155–162.

DebRoy S, Gebbie M, Ramesh A, Goodson JR, Cruz MR, van Hoof A, Winkler WC & Garsin DA (2014) '*A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator*', *Science*, Vol. **345**, pp. 937–940.

DeJesus MA, Sacchettini JC & Ioerger TR (2013) '*Reannotation of translational start sites in the genome of Mycobacterium tuberculosis*', *Tuberculosis*, Vol. **93**, pp. 18–25.

Demina GR, Nikitushkin VD, Shleeva MO, Riabova OB, Lepioshkin AY, Makarov VA & Kaprelyants AS (2017) '*Benzoylphenyl thiocyanates are new, effective inhibitors of the mycobacterial resuscitation promoting factor B protein*', *Annals of Clinical Microbiology and Antimicrobials*, Vol. **16**, p. 69.

Demirci H, Murphy F, Belardinelli R, Kelley AC, Ramakrishnan V, Gregory ST, Dahlberg AE & Jogl G (2010) '*Modification of 16S ribosomal RNA by the KsgA methyltransferase restructures the 30S subunit to optimize ribosome function*', *RNA*, Vol. **16**, pp. 2319–2324.

Downing KJ, Betts JC, Young DI, McAdam RA, Kelly F, Young M & Mizrahi V (2004) '*Global expression profiling of strains harbouring null mutations reveals that the five rpf-like genes of Mycobacterium tuberculosis show functional redundancy*', *Tuberculosis*, Vol. **84**, pp. 167–179.

Downing KJ, Mischenko V, Shleeva MO, Young DI, Young M, Kaprelyants AS, Apt AS & Mizrahi V (2005) '*Mutants of Mycobacterium tuberculosis lacking three of the five rpf-like genes are defective for growth in vivo and for resuscitation in vitro*', *Infection and Immunity*, Vol. **73**, pp. 3038–3043.

Frieda KL & Block SM (2012) '*Direct observation of cotranscriptional folding in an adenine riboswitch*', *Science*, Vol. **338**, pp. 397–400.

Galagan JE, Sisk P, Stolte C, Weiner B, Koehrsen M, Wymore F, Reddy TBK, Zucker JD, Engels R, Gellesch M, Hubble J, Jin H, Larson L, Mao M, Nitzberg M, White J, Zachariah ZK, Sherlock G, Ball CA & Schoolnik GK (2010) '*TB database 2010: overview and update*', *Tuberculosis*, Vol. **90**, pp. 225–235.

Gengenbacher M, Rao SPS, Pethe K & Dick T (2010) '*Nutrient-starved, non-replicating Mycobacterium tuberculosis requires respiration, ATP synthase and isocitrate lyase for maintenance of ATP homeostasis and viability*', *Microbiology*, Vol. **156**, pp. 81–87.

Gerhart E, Wagner H & Vogel J (2008) '*Approaches to identify novel non-messenger RNAs in bacteria and to investigate their biological functions: functional analysis of identified non-mRNAs*', *Handbook of RNA Biochemistry*, pp. 614–642.

Ghosh P, Ishihama A & Chatterji D (2001) '*Escherichia coli RNA polymerase subunit omega and its N-terminal domain bind full-length beta' to facilitate incorporation into the alpha(2)beta subassembly*', *European Journal of Biochemistry*, Vol. **268**, pp. 4621–4627.

'*Global tuberculosis report 2017*' (2017) World Health Organisation.

Gonzalez-y-Merchand JA, Garcia MJ, Gonzalez-Rico S, Colston MJ & Cox RA (1997) '*Strategies used by pathogenic and nonpathogenic mycobacteria to synthesize rRNA*', *Journal of Bacteriology*, Vol. **179**, pp. 6949–6958.

Gordon BRG (2013) '*Lsr2: an H-NS functional analog and global regulator of Mycobacterium tuberculosis*', *University of Toronto, Doctoral Thesis*.

Gordon BRG, Li Y, Wang L, Sintsova A, Bakel H van, Tian S, Navarre WW, Xia B & Liu J (2010) '*Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis*', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. **107**, pp. 5154–5159.

Gottesman S (2005) '*Micros for microbes: non-coding regulatory RNAs in bacteria*', *Trends in Genetics*, Vol. **21**, pp. 399–404.

Greive SJ & von Hippel PH (2005) '*Thinking quantitatively about transcriptional regulation*', *Nature Reviews Molecular Cell Biology*, Vol. **6**, pp. 221–232.

Grundy FJ, Winkler WC & Henkin TM (2002) '*tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome*', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. **99**, pp. 11121–11126.

Gupta RK, Srivastava BS & Srivastava R (2010) '*Comparative expression analysis of rpf-like genes of Mycobacterium tuberculosis H37Rv under different physiological stress and growth conditions*', *Microbiology*, Vol. **156**, pp. 2714–2722.

Gusarov I & Nudler E (1999) '*The mechanism of intrinsic transcription termination*', *Molecular Cell*, Vol. **3**, pp. 495–504.

Ha KS, Toulokhonov I, Vassylyev DG & Landick R (2010) '*The NusA N-terminal domain Is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase*', *Journal of Molecular Biology*, Vol. **401**, pp. 708–725.

Herrera-Asmat O, Lubkowska L, Kashlev M, Bustamante CJ, Guerra DG & Kireeva ML (2017) '*Production and characterization of a highly pure RNA polymerase holoenzyme from Mycobacterium tuberculosis*', *Protein Expression and Purification*, Vol. **134**, pp. 1–10.

Hett EC, Chao MC, Deng LL & Rubin EJ (2008) '*A mycobacterial enzyme essential for cell division synergizes with resuscitation-promoting factor*', *PLoS Pathogens*, Vol. **4**, p. e1000001.

Hett EC, Chao MC, Steyn AJ, Fortune SM, Deng LL & Rubin EJ (2007) '*A partner for the resuscitation-promoting factors of Mycobacterium tuberculosis*', *Molecular*

Microbiology, Vol. **66**, pp. 658–668.

Hett EC & Rubin EJ (2008) '*Bacterial growth and cell division: a mycobacterial perspective*', *Microbiology and Molecular Biology Reviews*, Vol. **72**, pp. 126–156.

Johansson J, Mandin P, Renzoni A, Chiaruttini C, Springer M & Cossart P (2002) '*An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes*', *Cell*, Vol. **110**, pp. 551–561.

Kaberdina AC, Szaflarski W, Nierhaus KH & Moll I (2009) '*An unexpected type of ribosomes induced by kasugamycin: a look into ancestral times of protein synthesis?*', *Molecular Cell*, Vol. **33**, pp. 227–236.

Kana BD, Gordhan BG, Downing KJ, Sung N, Vostroktunova G, Machowski EE, Tsenova L, Young M, Kaprelyants A, Kaplan G & Mizrahi V (2008) '*The resuscitation-promoting factors of Mycobacterium tuberculosis are required for virulence and resuscitation from dormancy but are collectively dispensable for growth in vitro*', *Molecular Microbiology*, Vol. **67**, pp. 672–684.

Kana BD, Karakousis PC, Parish T & Dick T (2014) '*Future target-based drug discovery for tuberculosis?*', *Tuberculosis*, Vol. **94**, pp. 551–556.

Kana BD & Mizrahi V (2010) '*Resuscitation-promoting factors as lytic enzymes for bacterial growth and signaling*', *FEMS Immunology and Medical Microbiology*, Vol. **58**, pp. 39–50.

Kapopoulou A, Lew JM & Cole ST (2011) '*The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes*', *Tuberculosis*, Vol. **91**, pp. 8–13.

Keep NH, Ward JM, Cohen-Gonsaud M & Henderson B (2006) '*Wake up! Peptidoglycan lysis and bacterial non-growth states*', *Trends in Microbiology*, Vol. **14**, pp. 271–276.

Kellenberger CA, Wilson SC, Hickey SF, Gonzalez TL, Su Y, Hallberg ZF, Brewer TF, Iavarone AT, Carlson HK, Hsieh Y-F & Hammond MC (2015) '*GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP*', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. **112**, pp. 5383–5388.

Kondratieva T, Rubakova E, Kana BD, Biketov S, Potapov V, Kaprelyants A & Apt A (2011) '*Mycobacterium tuberculosis attenuated by multiple deletions of rpf genes effectively protects mice against TB infection*', *Tuberculosis*, Vol. **91**, pp. 219–223.

Kyuma T, Kizaki H, Ryuno H, Sekimizu K & Kaito C (2015) '*16S rRNA methyltransferase KsgA contributes to oxidative stress resistance and virulence in Staphylococcus aureus*', *Biochimie*, Vol. **119**, pp. 166–174.

Landick R, Wang D & Chan CL (1996) '*Quantitative analysis of transcriptional pausing by Escherichia coli RNA polymerase: his leader pause site as paradigm*', *Methods in Enzymology*, Vol. **274**, pp. 334–353.

Lange C, Lehr M, Zerulla K, Ludwig P, Schweitzer J, Polen T, Wendisch VF & Soppa J (2017) '*Effects of Kasugamycin on the translatome of Escherichia coli*', *PLoS ONE*, Vol. **12**, p. e0168143.

Laursen B, Sørensen HP, Mortensen KK & Sperling-Petersen HU (2005) '*Initiation of protein synthesis in bacteria*', *Microbiology and Molecular Biology Reviews*, Vol. **69**, pp. 101–123.

Lew JM, Kapopoulou A, Jones LM & Cole ST (2011) '*TubercuList - 10 years after*', *Tuberculosis*, Vol. **91**, pp. 1–7.

Lillebaek T, Dirksen A, Baess I, Strunge B, Thomsen VØ & Andersen AB (2002) '*Molecular evidence of endogenous reactivation of Mycobacterium tuberculosis after 33 years of latent infection*', *The Journal of Infectious Diseases*, Vol. **185**, pp. 401–404.

Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S, Feng Y, Zhang Y, Mandal S, Jiang Y, Liu S, Gigliotti M, Talaue M, Connel N, Das K, Arnold E & Ebright RH (2017) '*Structural basis of Mycobacterium tuberculosis transcription and transcription inhibition structures of Mycobacterium tuberculosis RNA polymerase reveal taxon-specific properties and binding sites of known and new antituberculosis agents*', *Molecular Cell*, Vol. **66**, pp. 169–179.

Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, Repoila F, Buchrieser C, Cossart P & Johansson J (2009) '*A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes*', *Cell*, Vol. **139**, pp. 770–779.

Lutz B, Faber M, Verma A, Klumpp S & Schug A (2014) '*Differences between cotranscriptional and free riboswitch folding*', *Nucleic Acids Research*, Vol. **42**, pp. 2687–2696.

Mangat CS & Brown ED (2008) '*Ribosome biogenesis; the KsgA protein throws a methyl-mediated switch in ribosome assembly*', *Molecular Microbiology*, Vol. **70**, pp. 1051–1053.

Marr MT & Roberts JW (2000) '*Function of transcription cleavage factors GreA and GreB at a regulatory pause site*', *Molecular Cell*, Vol. **6**, pp. 1275–1285.

Mathew R, Mukherjee R, Balachandar R & Chatterji D (2005) '*Deletion of the gene rpoZ, encoding the omega subunit of RNA polymerase, in Mycobacterium smegmatis results in fragmentation of the beta' subunit in the enzyme assembly*', *Journal of Bacteriology*, Vol. **187**, pp. 6565–6570.

McCown PJ, Corbino KA, Stav S, Sherlock ME & Breaker RR (2017) '*Riboswitch diversity and distribution*', *RNA*, Vol. **23**, pp. 995–1011.

Mellin JR & Cossart P (2015) '*Unexpected versatility in bacterial riboswitches*', *Trends in Genetics*, Vol. **31**, pp. 150–156.

Mellin JR, Koutero M, Dar D, Nahori M-A, Sorek R & Cossart P (2014) '*Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA*', *Science*, Vol. **345**, pp. 940–943.

Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N & Breaker RR (2011) '*Challenges of ligand identification for riboswitch candidates*', *RNA Biology*, Vol. **8**, pp. 5–10.

Miah K (2014) '*Novel infection model of Mycobacteria, and characterising ribo-regulators of mycobacterium tuberculosis RpfB - its impact on pathogenesis and reactivation*' University College London.

Miller JH (1972) '*Experiments in molecular genetics*', in *Experiments in molecular genetics*, Cold Spring Harbor Laboratory Press, pp. 352–355.

Millman A, Dar D, Shamir M & Sorek R (2017) '*Computational prediction of regulatory, premature transcription termination in bacteria*', *Nucleic Acids Research*, Vol. **45**, pp. 886–893.

Montange RK & Batey RT (2008) '*Riboswitches: emerging themes in RNA structure and function*', *Annual Review of Biophysics*, Vol. **37**, pp. 117–133.

Moores A, Riesco AB, Schwenk S & Arnvig KB (2017) '*Expression, maturation and turnover of DrrS, an unusually stable, DosR regulated small RNA in Mycobacterium tuberculosis*', *PLoS ONE*, Vol. **12**, p. e0174079.

Mukamolova G, Kaprelyants A, Kell DB & Young M (2003) '*Adoption of the transiently non-culturable state - a bacterial survival strategy?*', *Advances in Microbial Physiology*, Vol. **47**, pp. 65–129.

Mukamolova G, Kaprelyants AS, Young DI, Young M & Kell DB (1998) '*A bacterial cytokine*', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. **95**, pp. 8916–8921.

Mukamolova G, Murzin AG, Salina EG, Demina GR, Kell DB, Kaprelyants AS & Young M (2006) '*Muralytic activity of Micrococcus luteus Rpf and its relationship to physiological activity in promoting bacterial growth and resuscitation*', *Molecular Microbiology*, Vol. **59**, pp. 84–98.

Mukamolova G, Turapov O, Kazarian K, Telkov M, Kaprelyants AS, Kell DB & Young M (2002) '*The rpf gene of Micrococcus luteus encodes an essential secreted growth factor*', *Molecular Microbiology*, Vol. **46**, pp. 611–621.

Mukamolova G, Turapov O, Young DI, Kaprelyants AS, Kell DB & Young M (2002) '*A family of autocrine growth factors in Mycobacterium tuberculosis*', *Molecular Microbiology*, Vol. **46**, pp. 623–635.

Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL & Breaker RR (2002) '*Genetic control by a metabolite binding mRNA*', *Chemistry and Biology*, Vol. **9**, pp. 1043–1049.

Nechooshtan G, Elgrably-Weiss M, Sheaffer A, Westhof E & Altuvia S (2009) '*A pH-responsive riboregulator*', *Genes and Development*, Vol. **23**, pp. 2650–2662.

Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX & Breaker RR (2013) '*Riboswitches in eubacteria sense the second messenger cyclic di-AMP*', *Nature Chemical Biology*, Vol. **9**, pp. 834–839.

Newton-Foot M & Gey Van Pittius NC (2013) '*The complex architecture of mycobacterial promoters*', *Tuberculosis*, Vol. **93**, pp. 60–74.

Nikitushkin VD, Demina GR & Kaprelyants AS (2016) '*Rpf proteins are the factors of reactivation of the dormant forms of actinobacteria*', *Biochemistry*, Vol. **81**, pp. 1719–1734.

Nikitushkin VD, Demina GR, Shleeva MO, Guryanova S, Ruggiero A, Berisio R & Kaprelyants AS (2015) '*A product of RpfB and RipA joint enzymatic action promotes the resuscitation of dormant mycobacteria*', *FEBS Journal*, Vol. **282**, pp. 2500–2511.

Nudler E & Mironov AS (2004) '*The riboswitch control of bacterial metabolism*', *Trends in Biochemical Sciences*, Vol. **29**, pp. 11–17.

Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR & Hatfull GF (2005) '*GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria*', *Cell*, Vol. **123**, pp. 861–873.

Ojha AK, Baughn AD, Sambandan D, Hsu T, Trivelli X, Guerardel Y, Alahari A, Kremer L, Jacobs WR & Hatfull GF (2008) '*Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria*', *Molecular Microbiology*, Vol. **69**, pp. 164–174.

Ojha AK & Hatfull GF (2012) '*Biofilms of Mycobacterium tuberculosis: new perspectives of an old pathogen*', in *Understanding tuberculosis - deciphering the secret life of the bacilli*, Intech, pp. 181–192.

Oliver JD (2005) '*The viable but nonculturable state in bacteria*', *The Journal of Microbiology*, Vol. **43**, pp. 93–100.

Papenfort K & Vogel J (2010) '*Regulatory RNA in bacterial pathogens*', *Cell Host & Microbe*, Vol. **8**, pp. 116–127.

Parveen N & Cornell KA (2011) '*Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism*', *Molecular Microbiology*, Vol. **79**, pp. 7–20.

Pedrolli D, Langer S, Hobl B, Schwarz J, Hashimoto M & Mack M (2015) '*The ribB FMN riboswitch from Escherichia coli operates at the transcriptional and translational level and regulates riboflavin biosynthesis*', *FEBS Journal*, Vol. **282**, pp. 3230–3242.

Perdrizet GA, Artsimovitch I, Furman R, Sosnick TR & Pan T (2012) '*Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch*', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. **109**, pp. 3323–3328.

Peters JM, Vangeloff AD & Landick R (2011) '*Bacterial transcription terminators: the RNA 3'-end chronicles*', *Journal of Molecular Biology*, Vol. **412**, pp. 793–813.

Phillips DR, Cutts SM, Cullinane CM & Crothers DM (2001) '*High-resolution transcription assay for probing drug-DNA interactions at individual drug sites*', *Methods in Enzymology*, Vol. **340**, pp. 466–485.

du Plessis J, Cloete R, Burchell L, Sarkar P, Warren RM, Christoffels A, Wigneshweraraj S & Sampson SL (2017) '*Exploring the potential of T7 bacteriophage protein Gp2 as a novel inhibitor of mycobacterial RNA polymerase*', *Tuberculosis*, Vol. **106**, pp. 82–90.

Polaski JT, Webster SM, Johnson JE & Batey RT (2017) '*Cobalamin riboswitches exhibit a broad range of ability to discriminate between methylcobalamin and adenosylcobalamin*', *Journal of Biological Chemistry*, Vol. **292**, pp. 11650–11658.

Price IR, Grigg JC & Ke A (2014) '*Common themes and differences in SAM recognition among SAM riboswitches*', *Biochimica et Biophysica Acta*, Vol. **1839**, pp. 931–938.

Ravagnani A, Finan CL & Young M (2005) '*A novel firmicute protein family related to the actinobacterial resuscitation-promoting factors by non-orthologous domain displacement*', *BMC Genomics*, Vol. **6**, p. 39.

Ray-Soni A, Bellecourt MJ & Landick R (2016) '*Mechanisms of bacterial transcription termination: all good things must end*', *Annual Review of Biochemistry*, Vol. **85**, pp. 319–347.

Richards JP & Ojha AK (2014) '*Mycobacterial biofilms*', *Microbiology Spectrum*, Vol. **2**, p. MGM2-0004-2013.

Richardson JP (2002) '*Rho-dependent termination and ATPases in transcript termination*', *Biochimica et Biophysica Acta*, Vol. **1577**, pp. 251–260.

Rickman L, Scott C, Hunt DM, Hutchinson T, Menéndez MC, Whalan R, Hinds J, Colston MJ, Green J & Buxton RS (2005) '*A member of the cAMP receptor protein family of transcription regulators in Mycobacterium tuberculosis is required for virulence in mice and controls transcription of the rpfA gene coding for a resuscitation promoting factor*', *Molecular Microbiology*, Vol. **56**, pp. 1274–1286.

Rosser A, Stover C, Pareek M & Mukamolova G (2017) '*Resuscitation-promoting factors are important determinants of the pathophysiology in Mycobacterium tuberculosis infection*', *Critical Reviews in Microbiology*, Vol. **43**, pp. 621–630.

Roth A & Breaker RR (2009) '*The structural and functional diversity of metabolite-binding riboswitches*', *Annual Review of Biochemistry*, Vol. **78**, pp. 305–334.

Ruff EF, Record MT & Artsimovitch I (2015) '*Initial events in bacterial transcription initiation*', *Biomolecules*, Vol. **5**, pp. 1035–1062.

Ruggiero A, Tizzano B, Pedone E, Pedone C, Wilmanns M & Berisio R (2009) '*Crystal structure of the resuscitation-promoting factor ΔDUFRpfB from M. tuberculosis*', *Journal of Molecular Biology*, Vol. **385**, pp. 153–162.

Sachdeva P, Misra R, Tyagi AK & Singh Y (2010) '*The sigma factors of Mycobacterium tuberculosis: regulation of the regulators*', *FEBS Journal*, Vol. **277**, pp. 605–626.

Schwenk S, Moores A, Nobeli I, McHugh TD & Arnvig KB (2018) '*Cell-wall synthesis and ribosome maturation are coregulated by an RNA switch in Mycobacterium tuberculosis*', *Nucleic Acids Research*, accepted.

Seidi K & Jahanban-Esfahlan R (2013) '*A novel approach to eradicate latent TB: based on resuscitation promoting factors*', *Journal of Medical Hypotheses and Ideas*, Vol. **7**, pp. 69–74.

Serganov A (2010) '*Determination of riboswitch structures: light at the end of the tunnel?*', *RNA Biology*, Vol. **7**, pp. 98–103.

Sharma AK, Chatterjee A, Gupta S, Banerjee R, Mandal S, Mukhopadhyay J, Basu J & Kundu M (2015) '*MtrA, an essential response regulator of the MtrAB two-component system, regulates the transcription of resuscitation-promoting factor B of Mycobacterium tuberculosis*', *Microbiology*, Vol. **161**, pp. 1271–1281.

Shine J & Dalgarno L (1975) '*Determinant of cistron specificity in bacterial ribosomes*', *Nature*, Vol. **254**, pp. 34–38.

Smith AM, Fuchs RT, Grundy FJ & Henkin TM (2010) '*The SAM-responsive SMK box is a reversible riboswitch*', *Molecular Microbiology*, Vol. **78**, pp. 1393–1402.

Smollett KL, Fivian-Hughes AS, Smith JE, Chang A, Rao T & Davis EO (2009) '*Experimental*

determination of translational start sites resolves uncertainties in genomic open reading frame predictions - application to Mycobacterium tuberculosis', *Microbiology*, Vol. **155**, pp. 186–197.

Snapper SB, Melton RE, Mustafa S, Kieser T & Jacobs WR (1990) '*Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis*', *Molecular Microbiology*, Vol. **4**, pp. 1911–1919.

St-Onge RJ, Haiser HJ, Yousef MR, Sherwood E, Tschowri N, Al-Bassam M & Elliot MA (2015) '*Nucleotide second messenger-mediated regulation of a muralytic enzyme in Streptomyces*', *Molecular Microbiology*, Vol. **96**, pp. 779–795.

Steinert H, Sochor F, Wacker A, Buck J, Helmling C, Hiller F, Keyhani S, Noeske J, Grimm SK, Rudolph MM, Keller H, Mooney RA, Landick R, Suess B, Fürtig B, Wöhnert J & Schwalbe H (2017) '*Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation*', *eLife*, Vol. **6**, p. e21297.

Storz G, Opdyke JA & Zhang A (2004) '*Controlling mRNA stability and translation with small, noncoding RNAs*', *Current Opinion in Microbiology*, Vol. **7**, pp. 140–144.

Sudarsan N, Lee ER, Moy RH, Kim JN, Link KH & Breaker RR (2008) '*Riboswitches in eubacteria sense the second messenger cyclic di-GMP*', *Science*, Vol. **321**, pp. 411–413.

Sudarsan N, Wickiser JK, Nakamura S, Ebert MS & Breaker RR (2003) '*An mRNA structure in bacteria that controls gene expression by binding lysine*', *Genes and Development*, Vol. **17**, pp. 2688–2697.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ & von Mering C (2017) '*The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible*', *Nucleic Acids Research*, Vol. **45**, pp. D362–D368.

Tetone LE, Friedman LJ, Osborne ML, Ravi H, Kyzer S, Stumper SK, Mooney RA, Landick R & Gelles J (2017) '*Dynamics of GreB-RNA polymerase interaction allow a proofreading accessory protein to patrol for transcription complexes needing rescue*', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. **114**, pp. E1081–E1090.

Thomas CM (2001) '*lacZ fusions report gene expression don't they?*', *Microbiology*, Vol. **147**, pp. 1993–1995.

Tucker BJ & Breaker RR (2005) '*Riboswitches as versatile gene control elements*', *Current Opinion in Structural Biology*, Vol. **15**, pp. 342–348.

Tufariello JM, Jacobs WR & Chan J (2004) '*Individual Mycobacterium tuberculosis resuscitation-promoting factor homologues are dispensable for growth in vitro and in vivo individual Mycobacterium tuberculosis*', *Infection and Immunity*, Vol. **72**, pp. 515–526.

Tufariello JM, Mi K, Xu J, Manabe YC, Kesavan AK, Drumm J, Tanaka K, Jacobs WR & Chan J (2006) '*Deletion of the Mycobacterium tuberculosis resuscitation-promoting factor Rv1009 gene results in delayed reactivation from chronic tuberculosis*', *Infection and Immunity*, Vol. **74**, pp. 2985–2995.

Umezawa H, Hamada M, Suhara Y, Hashimoto T & Ikekawa T (1965) '*Kasugamycin, a new*

antibiotic', *Antimicrobial Agents and Chemotherapy*, Vol. **5**, pp. 753–757.

Wagner EGH, Altuvia S & Romby P (2002) '*Antisense RNAs in bacteria and their genetic elements*', *Advances in Genetics*, Vol. **46**, pp. 361–398.

Wards BJ & Collins DM (1996) '*Electroporation at elevated temperatures substantially improves transformation efficiency of slow-growing mycobacteria*', *FEMS Microbiology Letters*, Vol. **145**, pp. 101–105.

Warner DF, Savvi S, Mizrahi V & Dawes SS (2007) '*A riboswitch regulates expression of the coenzyme B12-independent methionine synthase in Mycobacterium tuberculosis: implications for differential methionine synthase function in strains H37Rv and CDC1551*', *Journal of Bacteriology*, Vol. **189**, pp. 3655–3659.

Waters LS & Storz G (2009) '*Regulatory RNAs in bacteria*', *Cell*, Vol. **136**, pp. 615–628.

Weiss A, Moore BD, Tremblay MHJ, Chaput D, Kremer A & Shaw LN (2017) '*The omega subunit governs RNA polymerase stability and transcriptional specificity in Staphylococcus aureus*', *Journal of Bacteriology*, Vol. **199**, pp. e00459-16.

Wickiser JK, Winkler WC, Breaker RR & Crothers DM (2005) '*The speed of RNA transcription and metabolite dinding kinetics operate an FMN riboswitch*', *Molecular Cell*, Vol. **18**, pp. 49–60.

Winkler WC, Nahvi A, Roth A, Collins J & Breaker RR (2004) '*Control of gene expression by a natural metabolite-responsive ribozyme*', *Nature*, Vol. **428**, pp. 281–286.

Xu Z, O'Farrell HC, Rife JP & Culver GM (2008) '*A conserved rRNA methyltransferase regulates ribosome biogenesis*', *Nature Structural & Molecular Biology*, Vol. **15**, pp. 534–536.

Yamamoto H, Wittek D, Gupta R, Qin B, Ueda T, Krause R, Yamamoto K, Albrecht R, Pech M & Nierhaus KH (2016) '*70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria*', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. **113**, pp. E1180–E1189.

Zambrano MM & Kolter R (2005) '*Mycobacterial biofilms: a greasy way to hold it together*', *Cell*, Vol. **123**, pp. 762–764.

Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G & Gottesman S (2003) '*Global analysis of small RNA and mRNA targets of Hfq*', *Molecular Microbiology*, Vol. **50**, pp. 1111–1124.

Zhang J & Landick R (2016) '*A two-way street: regulatory interplay between RNA polymerase and nascent RNA structure*', *Trends in Biochemical Sciences*, Vol. **41**, pp. 293–310.

Zhonghe S & Zhang Y (1999) '*Spent culture supernatant of Mycobacterium tuberculosis H37Ra improves viability of aged cultures of this strain and allows small inocula to initiate growth*', *Journal of Bacteriology*, Vol. **181**, pp. 7626–7628.

Zuker M (2003) '*Mfold web server for nucleic acid folding and hybridization prediction*', *Nucleic Acids Research*, Vol. **31**, pp. 3406–3415.

10 Appendix

10.1 Constructs

Below is a detailed list of constructs employed in this study:

Construct	Source	Oligonucleotides
pGEM	pGEM T-easy II (Promega)	-
pCR4 TOPO	TOPO TA Cloning Kit with PCR4 TOPO (Invitrogen)	-
pKA425	(Moores *et al.* 2017)	-
pKA425p	pKA425 with PCLI heterologous promoter (Moores *et al.* 2017) inserted between XbaI and HindIII restriction	3.64 and 3.65
pEJ414-P1RpfB	Translational fusion of *rpfB* from -355 bp upstream to +1087 bp downstream of the annotated ATG start codon. Following +1087 bp downstream of ATG (Miah 2014)	-
425L-RpfB-ATG	pKA425 translational fusion of RpfB from -355 bp upstream of and including the annotated ATG start codon in fusion with *lacZ*	2.20 and 2.34
425L-TLSM-2	425L-RpfB-ATG with single basepair deletion +151 bp downstream of the P1 TSS	5.08 and 5.09
425L-TLSM-1	425L-RpfB-ATG with single basepair deletion +191 bp downstream of the P1 TSS	2.49 and 2.50
425L-TLSM+1	425L-RpfB-ATG with double basepair deletion in *lacZ* CDS at +139 to +140 bp downstream of the ATG start codon	2.64 and 2.65
425L-TLSM-GTc	425L-RpfB-ATG with GTG to GTc substitution +143 bp downstream of the P_1 TSS	5.17 with 2.40, and 5.16 with 2.73
425L-TLSM-TTa	425L-RpfB-ATG with TTG to TTa substitution +179 bp downstream of the P_1 TSS. Disrupts TTG start codon	5.19 with 2.40, and 5.18 with 2.73
425L-TLSM-AaG	425L-RpfB-ATG with ATG to AaG substitution +217 bp downstream of the P_1 TSS	2.70 and 2.71
425L-TLSM-RBS*	425L-RpfB-ATG with GAGGTCGGGGA to ctccTCcccct substitution +156 to +166 bp downstream of the P_1 TSS	2.74 and 2.75
425L-RpfB-RBSWterm	425L-RpfB-ATG with G to C substitution +112 bp downstream of the P_1 TSS	2.79 and 2.80
425L-RpfB-RBSWanti	425L-RpfB-ATG with U to C substitution +6 bp downstream of the P_1 TSS	5.02 and 5.03
425L-RpfB-RBSWdel	425L-RpfB-ATG with deletion spanning +1 bp to +127 bp of the P_1 TSS	5.37 and 5.38
425L-RpfB-RBSWpolA	425L-RpfB-ATG with U to A substitutions at +117 to +119 bp downstream of the P_1 TSS	5.33 with 2.40, and 5.32 with 2.73
425L-RpfB-P_1*P_2	425L-RpfB-ATG with TAGGGT to cAGGGc substitution -12 to -7 bp upstream of the P_1 TSS	8.37 with 2.40, and 8.36 with 2.73
425L-RpfB-P_1P_2*	425L-RpfB-ATG with TACCGT to cACCGc substitution +68 to +73 bp downstream of the P_1 TSS	8.39 with 2.40, and 8.38 with 2.73
425L-RpfB-P_1*P_2*	425L-RpfB-P1*P2 with TACCGT to cACCGc substitution +68 to +73 bp downstream of the P_1 TSS	8.39 with 2.40, and 8.38 with 2.73
425S-RpfB-asRNA100	pKA425 transcriptional fusion of *rpfB* asRNA spanning -100 bp upstream of the asRNA TSS to +2 bp	2.68 and 2.69
425P-KsgA-25bp	pKA425p translational fusion spanning -25 to +150 bp downstream of the *ksgA* ATG start codon, fused with *lacZ*	5.50 and 5.49
425P-KsgA-75bp	pKA425p translational fusion spanning -75 to +150 bp downstream of the *ksgA* ATG start codon, fused with *lacZ*	8.01 and 5.49
425P-KsgA-1120bp	pKA425p translational fusion spanning -1120 to +150 bp downstream of the *ksgA* ATG start codon, fused with *lacZ*	5.48 and 5.49

10.2 *In vitro* transcription template vectors

Construct	Characteristics	Oligonucleotides
pGAMrrnX_haltRBSW	pGEM containing *M. smegmatis* ribosomal *rrnB* P_1 promoter spanning -80 to -8 bp of the TSS (Cortes *et al.* 2013; Moores *et al.* 2017), inclusive of extended 'TGN' -10 motif. GTAAA insertion downstream if P_{rrnB} TSS. Transcriptional fusion of +1 to +176 of the *rpfB* P_1 TSS fused downstream of the GTAAA insertion. The entire template was inserted within T-A cloning site. Construction resulted in a 42 bp deletion of the pGEM backbone between the vector NdeI restriction site and T-A cloning site and was therefore renamed pGAM	5.59 and 5.60 (ext_rrnB), 5.15 and 5.23 (RBSW), 5.79 and 5.23 (halt)
pGAMrrnX_haltRBSWanti	pGAMrrnX_haltRBSW with U to C substitution at +11 bp downstream of P_{rrnB} TSS	8.03 and 5.23
pGAMrrnX_haltRBSWterm	pGAMrrnX_haltRBSW with G to C substitution at +117 bp downstream of P_{rrnB} TSS	5.79 and 5.23
pGAMrrnX_haltRBSWpolA	pGAMrrnX_haltRBSW with U to A substitutions at +122 to 124 bp downstream of P_{rrnB} TSS	5.79 and 5.23
pGAMrrnX_haltRBSW_STOP	pGAMrrnX_RBSW with *synB* synthetic terminator (Czyz *et al.* 2014) inserted +186 bp downstream of the P_{rrnB} TSS	8.44 and 8.45
pGAMrrnX_haltRBSWanti_STOP	pGAMrrnX_haltRBSW_STOP with U to C substitution at +11 bp downstream of P_{rrnB} TSS	8.44 and 8.45
pGAMrrnX_haltRBSWterm_STOP	pGAMrrnX_haltRBSW_STOP with G to C substitution at +117 bp downstream of P_{rrnB} TSS	8.44 and 8.45
pGAMrrnX_haltRBSWpolA_STOP	pGAMrrnX_haltRBSW_STOP with U to A substitutions at +122 to 124 bp downstream of P_{rrnB} TSS	8.44 and 8.45

10.3 Oligonucleotides

All oligonucleotides were commercially synthesised by Sigma, re-suspending the lyophilised stock in molecular grade H_2O to 100 µM.

Number	Sequence	Use
1.48	GTCCCATTCCGAACCCGGAAGCTAAGCCTGCCAGCGCCTGTCTC	Northern probe
2.07	GCTGTCAACGATACGCTACGTAACGGCATGACAGTGTTTTTTTTTTTTTTTTTTTTTTTT	3'RACE
2.09	GCTGTCAACGATACGCTACGTAACGGC	3'RACE
2.15	ATTGGCCGGTTCGTTACC	3'RACE
2.20	GCAGTCGATCGTACGCTAGT	Cloning
2.34	ATATATCCATGGGTGATTGGGTCTGATGAA	Cloning
2.40	ACGAGGGGCATTCACACCAGATTG	Cloning
2.49	CAAACTTCATCAGACCCAATC	Cloning
2.50	TAAGTAGAGTCAACGCGC	Cloning
2.64	CGTTTTACAACGTCGTGACTGGGAA	Cloning
2.65	CGGGATCATCCATGGGTGATTG	Cloning
2.68	TAATAAGCTTCGCGTTGACTCTACTTACCAAAC	Cloning
2.69	ATTATCTAGACATAGCCACCGGCGAACG	Cloning
2.70	CCAATCACCCAAGGATGATCCCG	Cloning
2.71	GTCTGATGAAGTTTGGTAAGTAGAG	Cloning
2.73	GCAGTTTGAGGGGACGACGACAGTATC	Cloning
2.74	CCCCTTATAGCGCGTTGACTCTAC	Cloning
2.75	GAGGAGGGCAACACTCAACCACCT	Cloning
2.79	GTTTCGGAGGCCCCATTTTGCTTTTTG	Cloning
2.80	CTAGAGGCCCCACCCGTT	Cloning
3.64	CTAGACGCTGACCACCCCCAGAACCCGAAGTTGACTCAAGTTCATTGGACTTGGTACAGTGA	Cloning
3.65	AGCTTCACTGTACCAAGTCCAATGAACTTGAGTCAACTTCGGGTTCTGGGGGTGGTCAGCGT	Cloning
5.02	GATGCGCCAACGAGCGCGCCG	Cloning
5.03	CACCCTAGCCCATAAGCTCGG	Cloning
5.08	TTGCCGAGGTCGGGGATATAGC	Cloning
5.09	ACTCAACCACCTACACAGCGAAC	Cloning
5.15	ATTAACTAGTGCCAATGAGCGCGCCG	Cloning
5.16	GTTCGCTGTGTAGGTCGTTGAGTGTT	Cloning
5.17	AACACTCAACGACCTACACAGCGAAC	Cloning
5.18	GGGGATATAGCGCGTTAACTCTACTTACC	Cloning
5.19	GGTAAGTAGAGTTAACGCGCTATATCCCC	Cloning
5.22	CGCGCCGAGCGGCCCATAACACCCGCGCGCCGGAGTTGCTCCTGTCTC	Northern probe
5.23	ATTACTCGAGCGCGCTATATCCCCGACC	Cloning/IVT template
5.31	CATATGGGTGACCGCGTCTGA	Cloning/IVT template
5.32	GGCCCAAAATGCTTTTTGTTCGCTG	Cloning

Number	Sequence	Use
5.33	CAGCGAACAAAAAGCATTTTGGGCC	Cloning
5.37	GTTCGCTGTGTAGGTGGTTGAGTGTTGC	Cloning
5.38	GCATCCACCCTAGCCCATAAGCTCG	Cloning
5.48	ATTAAAGCTTGAGGTCGGGGATATAGCGCGTTG	Cloning
5.49	TAATCCATGGCAACCACCCGTCGCACCGTG	Cloning
5.50	ATTAAAGCTTGTCAAGGTTGGGGCGCCTGG	Cloning
5.51	AACGGCGGGCTGCGGTATGC	Cloning
5.59	TATGGGTGACCGCGTCTGACCAGGGAAAATAGCCCTCTGACCTGGGGATTTGACTCCCAGTTTCC AAGGTGGTAACTTA	Cloning
5.60	CTAGTAAGTTACCACCTTGGAAACTGGGAGTCAAATCCCCAGGTCAGAGGGCTATTTTCCCTGGTC AGACGCGGTCACCCA	Cloning
5.79	ACTTACTAGTGTAAAGCCAATGAGCGCGCCGAGCGGCCCA	Cloning
8.01	ATTAAAGCTTCCTCGCCACCCGCGAAGAGC	Cloning
8.03	ACTTACTAGTGTAAAGCCAACGAGCGCGCCGAGCGGCCCA	Cloning
8.04	GTCTTGCGCCGCTGCGCGAA	RT-qPCR
8.06	CGCGCTATATCCCCGACCTCG	IVT template amplification
8.07	GGACCCAGGATGTGACGTTC	RT-qPCR
8.08	TGCACACCACCGTAATACCC	RT-qPCR
8.09	ATCTGCTTGTCGAGTTCCCG	RT-qPCR
8.10	AGGGCGAGGTCTCATATCGA	RT-qPCR
8.11	GAGCGGCCCATAACACCC	RT-qPCR
8.27	CGGCAACACTCAACCACCTACACAG	IVT template amplification
8.36	CTTATGGGCCAGGGCGGATGCGCCA	Cloning
8.37	TGGCGCATCCGCCCTGGCCCATAAG	Cloning
8.38	GCCGGTTCGTCACCGCCTTGTGATC	Cloning
8.39	GATCACAAGGCGGTGACGAACCGGC	Cloning
8.43	CCGCGTTGAAAAAAAAAAGCGCC	IVT template amplification
8.44	TCGAGCGCCGCAACTGCGGCGCTTTTTTTTTTCAACGCGGATCCTAATAATCG	Cloning
8.45	AATTCGATTATTAGGATCCGCGTTGAAAAAAAAAAGCGCCGCAGTTGCGGCGC	Cloning
8.53	CGTTCGATCACAAGACGGTA	RT-qPCR
8.56	TCAACATTGGCCGGTTCGTT	RT-qPCR
8.57	GACTACCAGGCGCAACATCG	RT-qPCR
8.58	TTATGGGCTAGGGTGGATGC	RT-qPCR

10.4 Supplemental figures

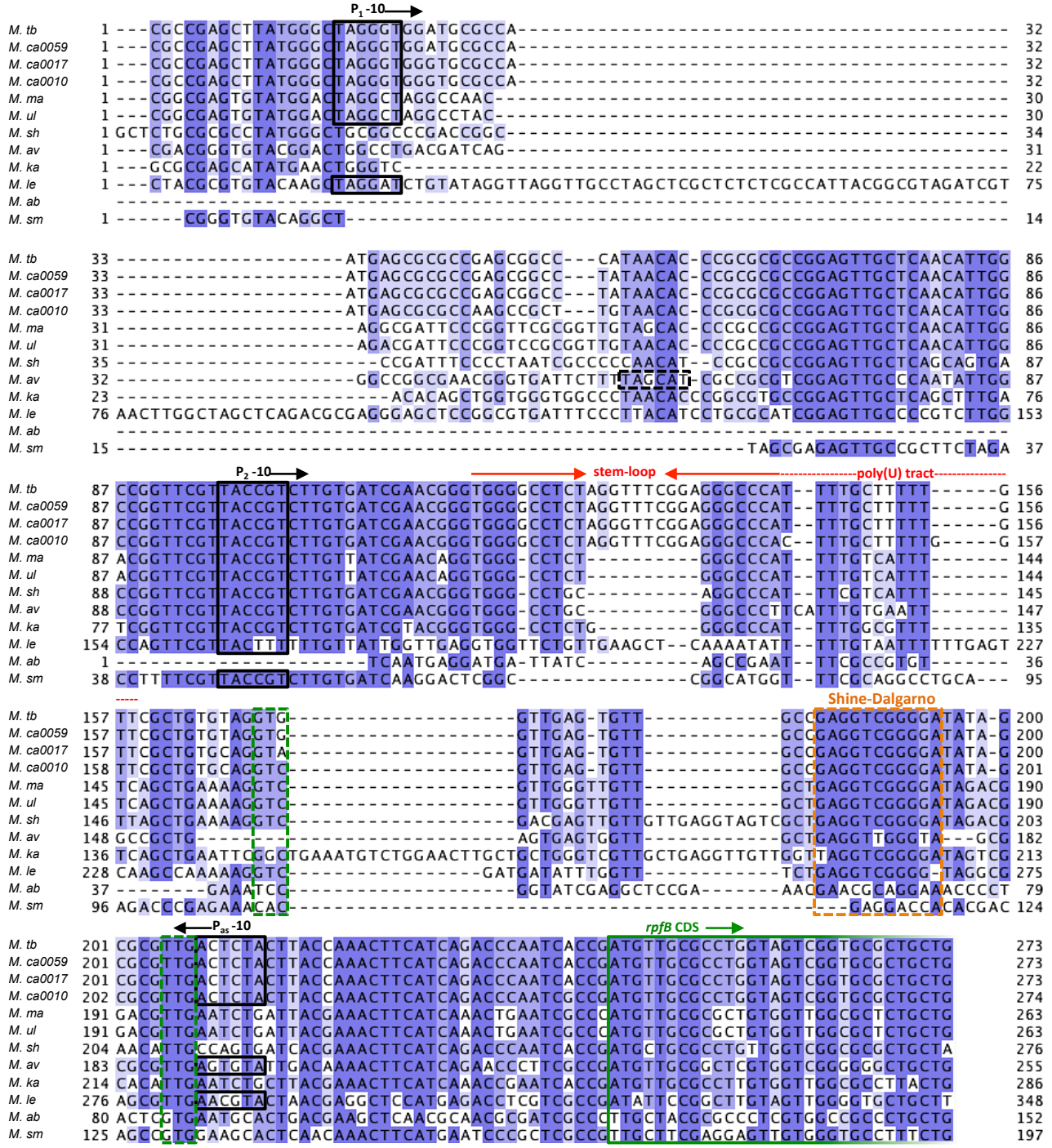


Figure 10.I extended sequence alignment of *rpfB* promoter regions and 5'UTR

Sequence alignment of *M. tuberculosis* (*M. tb*), three strains of *M. canettii*: CIPT 140010059 (*M. ca0059*), CIPT 140070017 (*M. ca0017*) and CIPT 140070010 (*M. ca0010*), *Mycobacterium marinum* (*M. ma*), *Mycobacterium ulcerans* (*M. ul*), *Mycobacterium shinjukuense* (*M. sh*), *Mycobacterium avium* (*M. av*), *Mycobacterium kansasii* (*M. ka*), *Mycobacterium leprae* (*M. le*), *Mycobacterium abscessus* (*M. ab*) and *Mycobacterium smegmatis* (*M. sm*) constructed in Jalview. Adapted from (Schwenk *et al.* 2018). Residues coloured by percentage sequence conservation (darker blue indicates higher conservation). Putative P_1, P_2 and P_{as} -10 promoter elements are highlighted in black. First ten codons of *rpfB* CDS highlighted in green, with alternative start codons highlighted with dashed lines. Proposed SD sequence is highlighted in orange (see section 3.4). Red annotations denote the stem-loop poly(U) tract of the proposed riboswitch intrinsic terminator identified in *M. tuberculosis*.

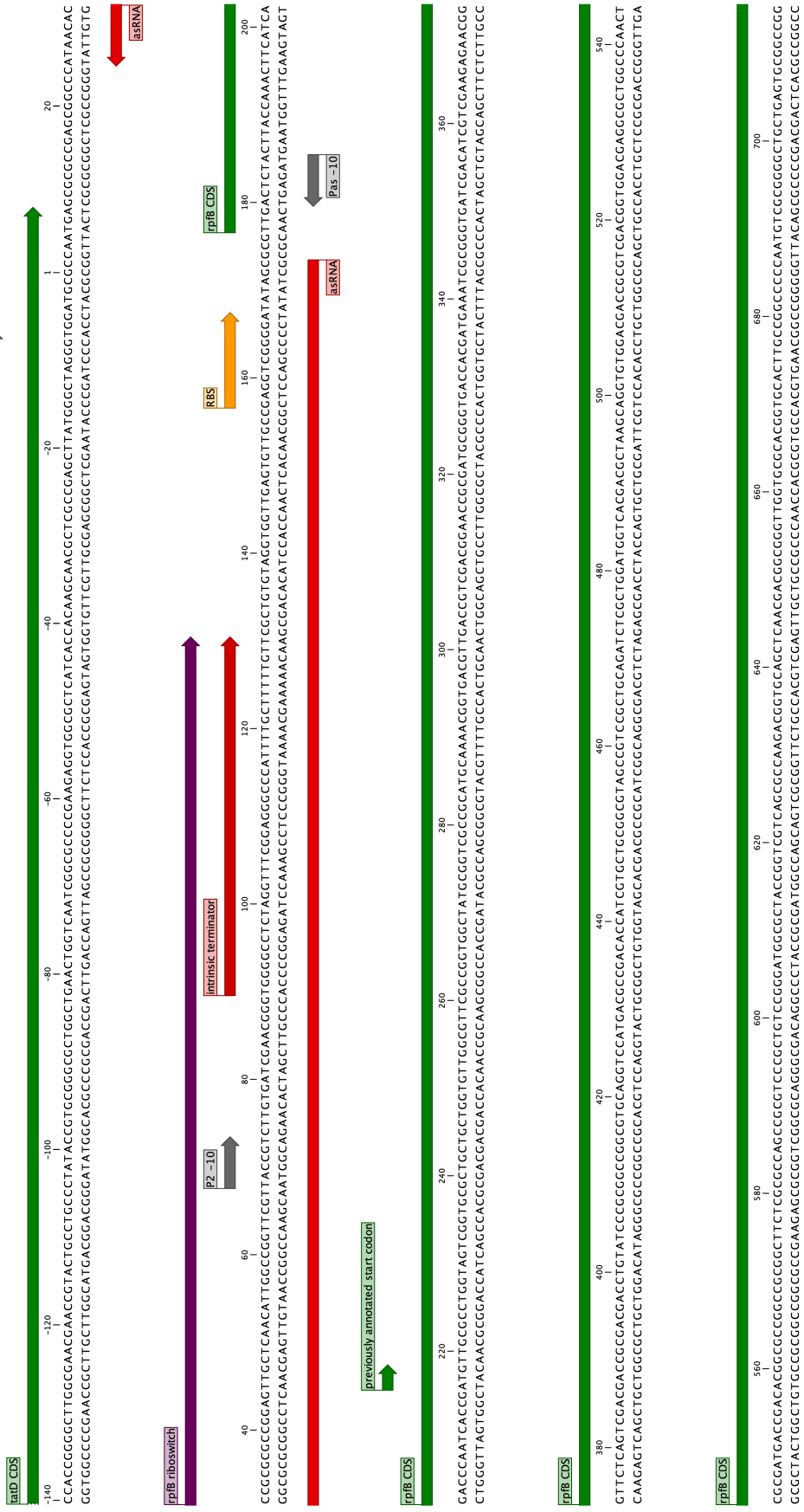


Figure 10.II full sequence of the *M. tuberculosis* H37RV *rpfB* locus
Positions 1127766 to 1130129 shown, experimentally validated *rpfB* TTG start codon annotated (see seciton 3).

Figure 10.II continued

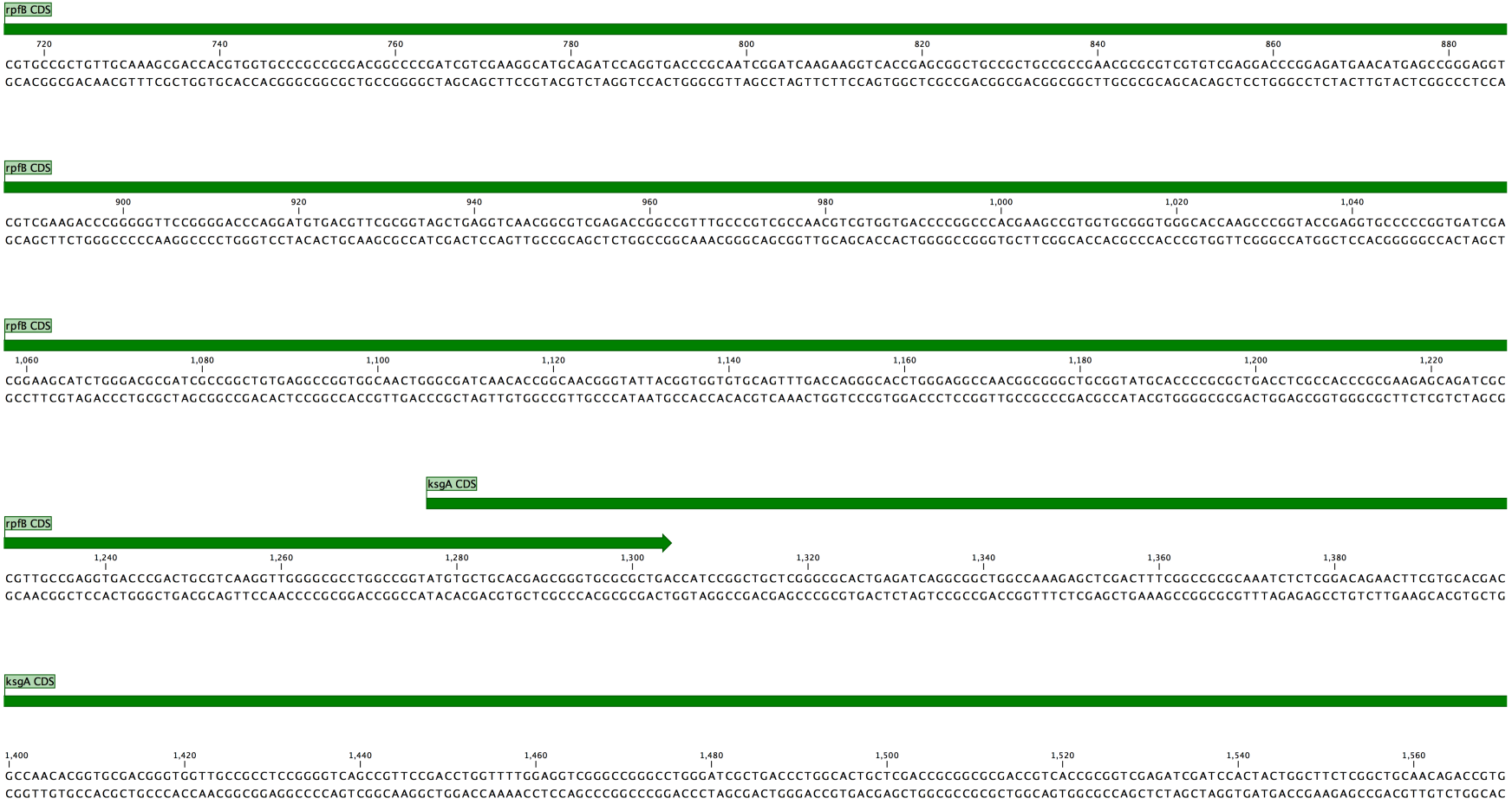


Figure 10.II continued

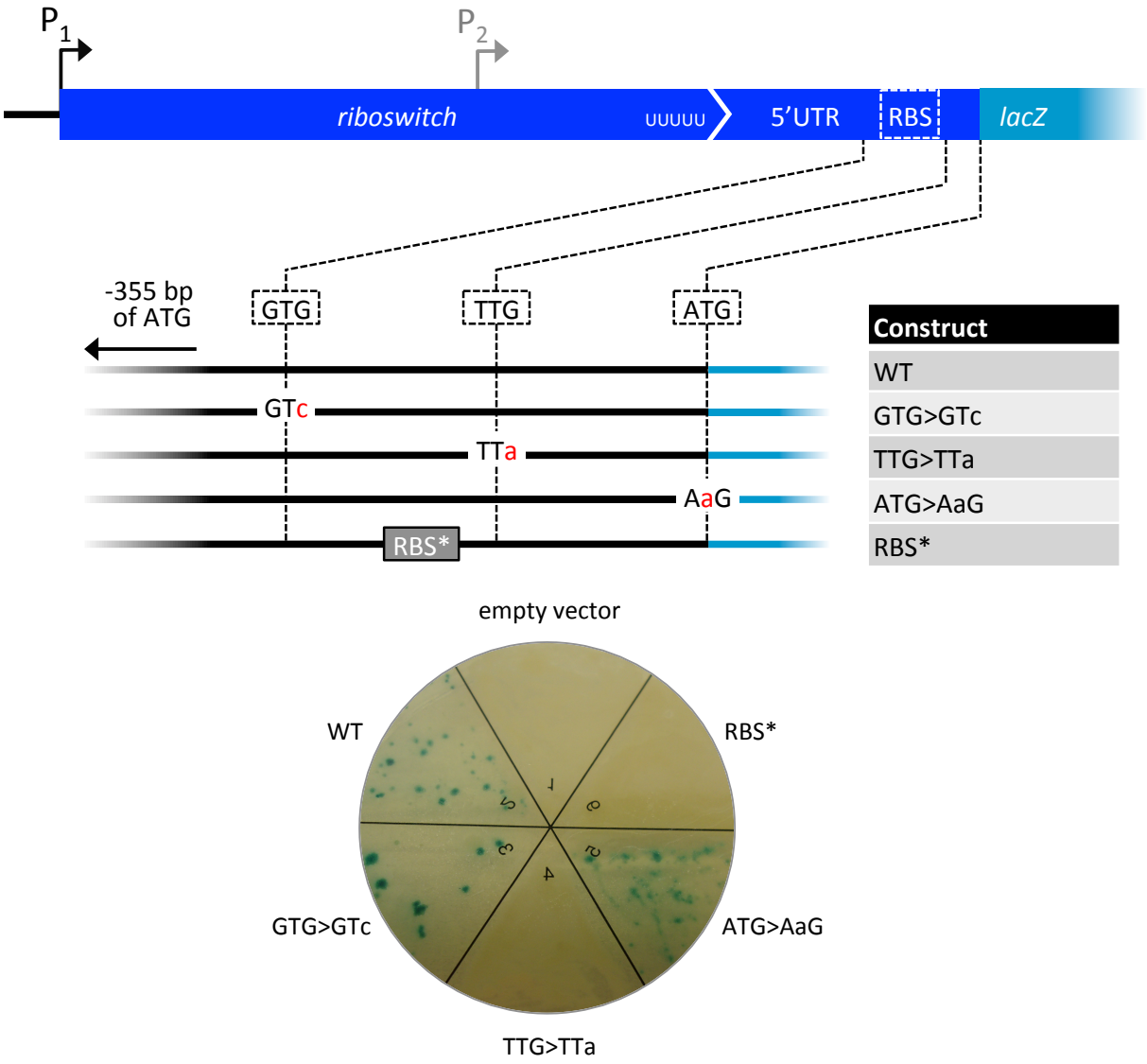


Figure 10.III translational reporters in *M. tuberculosis*
Top diagram depicts the mutations introduced to each construct, where red bases indicate substitutions and 'RBS*' indicates ribosome binding site mutation. Constructs match those described in section 3.3. Bottom image shows β-galactosidase expression in transformed *M. tuberculosis* on 7H11 agar.

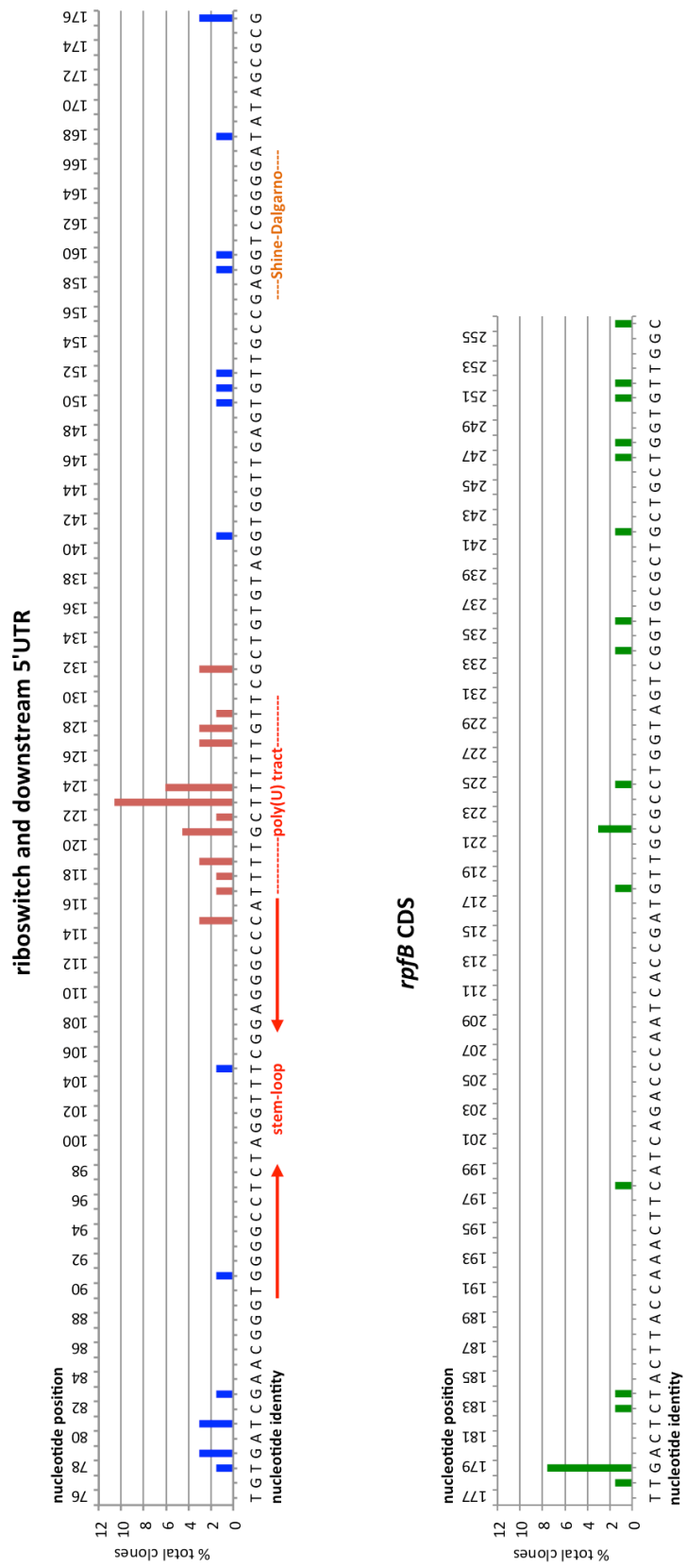


Figure 10.IV extended 3'RACE mapping
Chart showing mapped 3'termini of *rpfB* transcripts from *M. tuberculosis* H37Rv exponential growth phase total RNA. Total 3'termini mapped at single nucleotide resolution within the *rpfB* locus from +77 to +256 bp relative to the P_1 TSS.

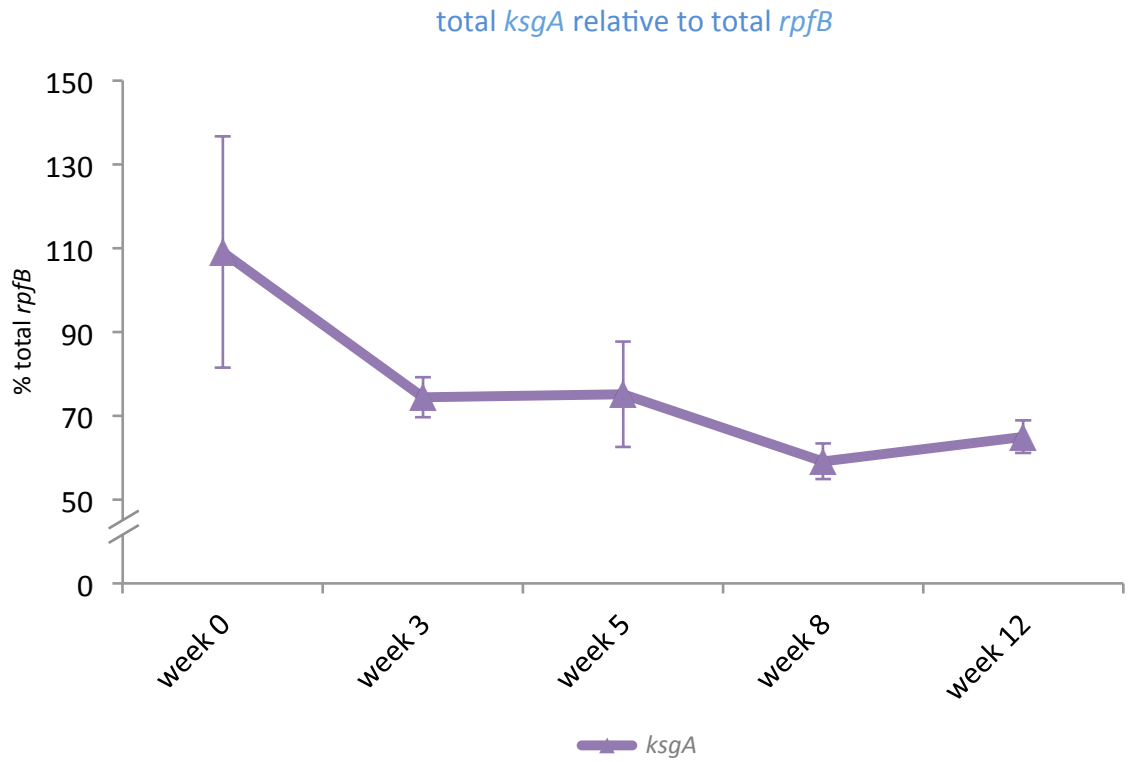


Figure 10.V *M. bovis* BCG biofilm RT-qPCR detection of *ksgA* relative to *rpfB*
Detected *ksgA* amplicons of *M. bovis* BCG biofilm cDNA expressed as a percentage of detected total *rpfB* amplicons. Data represents mean of technical replicates (n=4) of the biological replicates (n=3) ± standard deviation.

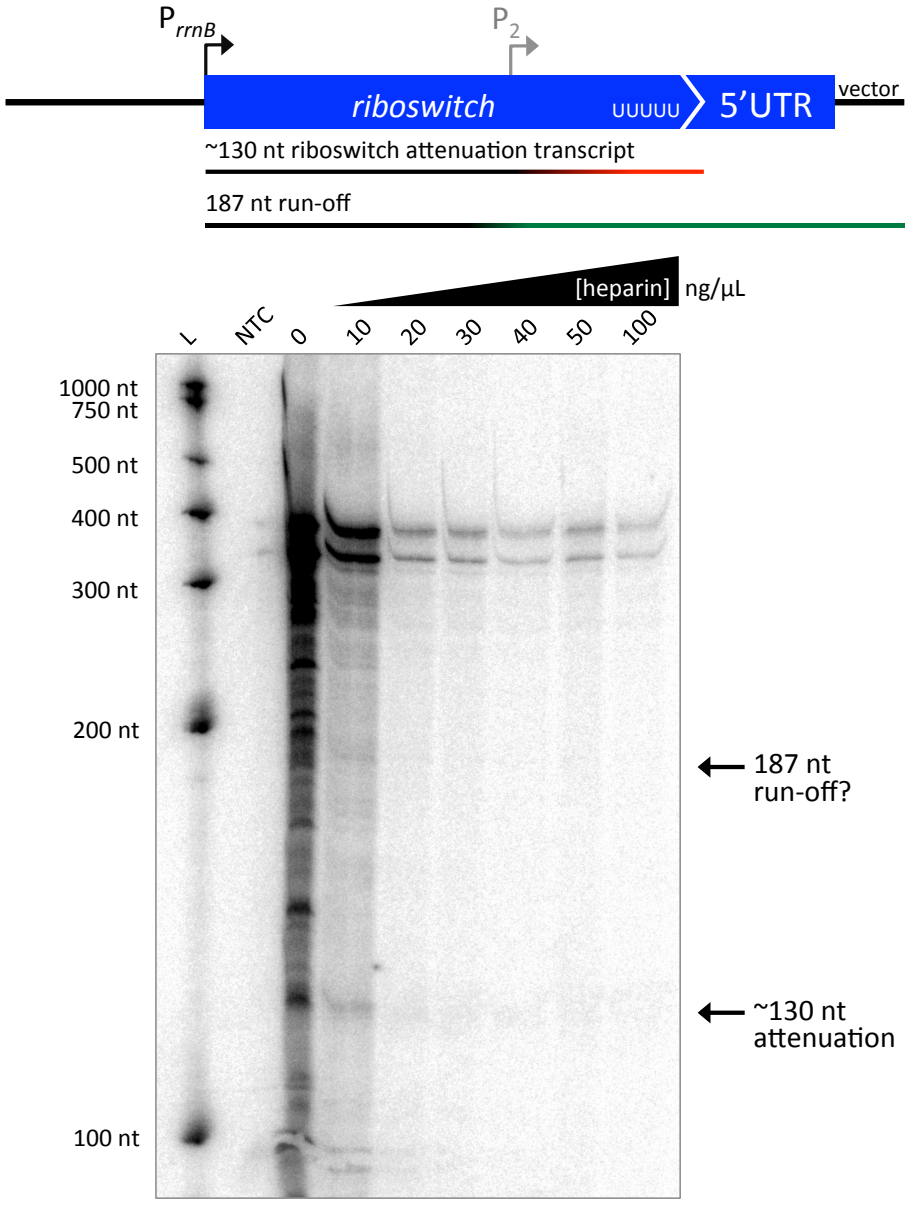


Figure 10.VI preliminary *in vitro* transcription with template digest
PAGE of multi-round *in vitro* transcription reactions. WT *M. smegmatis* P_{rrnB} promoted transcription of the riboswitch, with titration of heparin from 0 to 100 ng/µL. Reactions were carried out at 37°C with 17 nM *E. coli* RNAP (NEB), 200 nM template and 500 µM NTP, labelling with 32P-αUTP (Perkin Elmer). Template amplified with oligonucleotides 5.23 and 2.39 (see Appendix section 10.3).

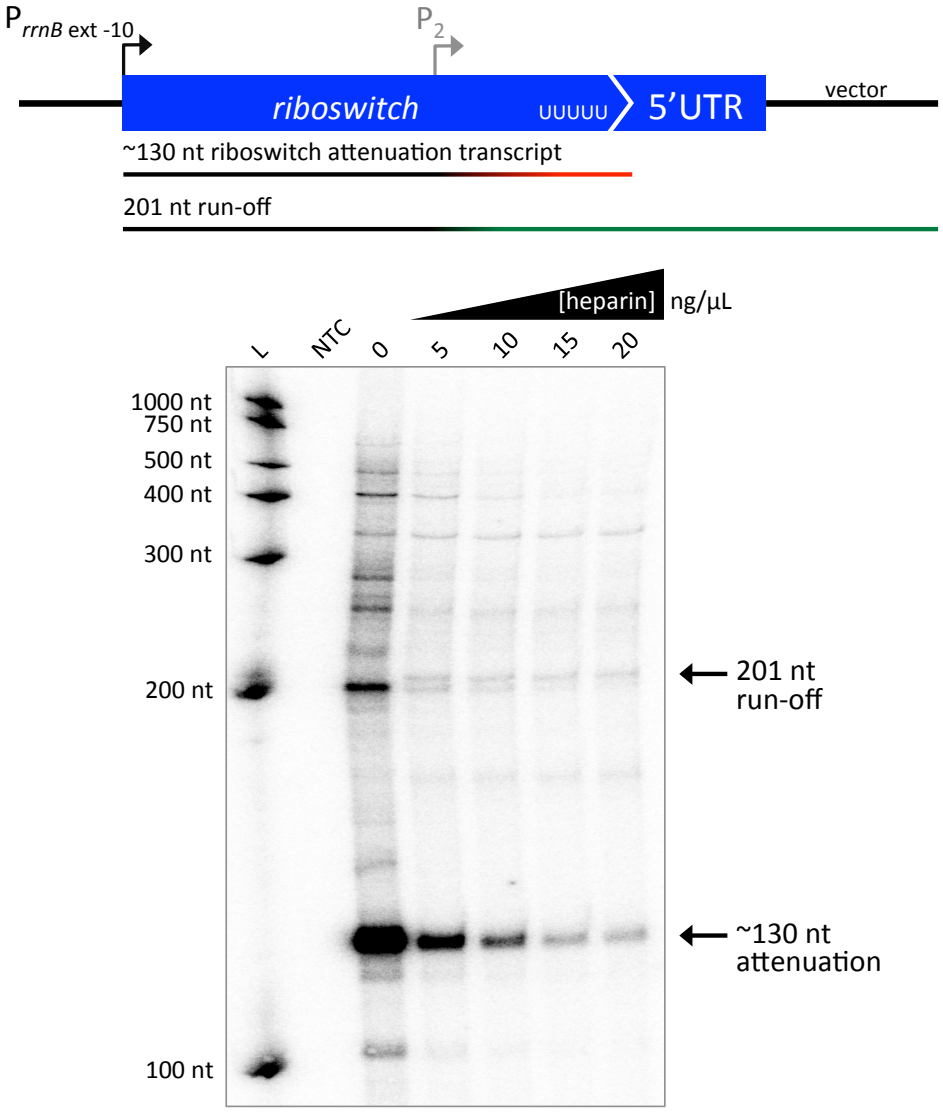


Figure 10.VII preliminary *in vitro* transcription with PCR template
PAGE of multi-round *in vitro* transcription reactions. Template with added G:C basepair, *M. smegmatis* extended -10 P_{rrnB} promoted transcription of the riboswitch, with titration of heparin from 0 to 20 ng/µL. Reactions were carried out at 37°C with 17 nM *E. coli* RNAP (NEB), 200 nM template and 500 µM NTP, labelling with 32P-αUTP (Perkin Elmer). Reactions were stopped after 20 min with the addition of formamide loading dye. Template amplified with oligonucleotides 5.61 and 5.62 (see Appendix section 10.3).

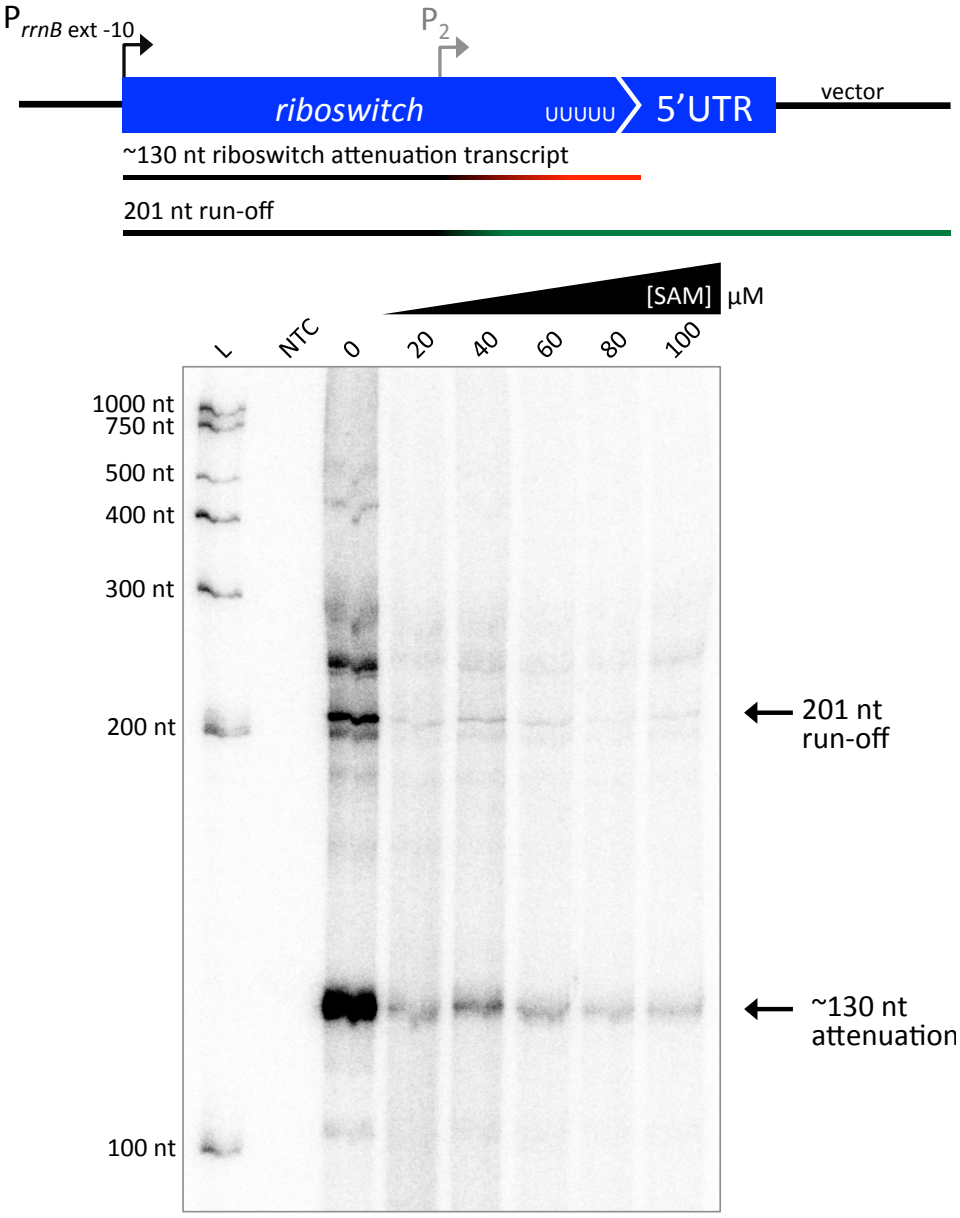


Figure 10.VIII preliminary *in vitro* transcription with candidate ligand
Multi-round *in vitro* transcription of the riboswitch in the presence of S-adenosyl-L-methionine titration from 0 to 100 µM. Reactions were carried out at 37°C with 17 nM *E. coli* RNAP (NEB), 200 nM template and 500 µM NTP, labelling with 32P-αUTP (Perkin Elmer). Reactions were stopping after 20 min with the addition of formamide loading dye. Template amplified with oligonucleotides 5.61 and 5.62 (see Appendix section 10.3).

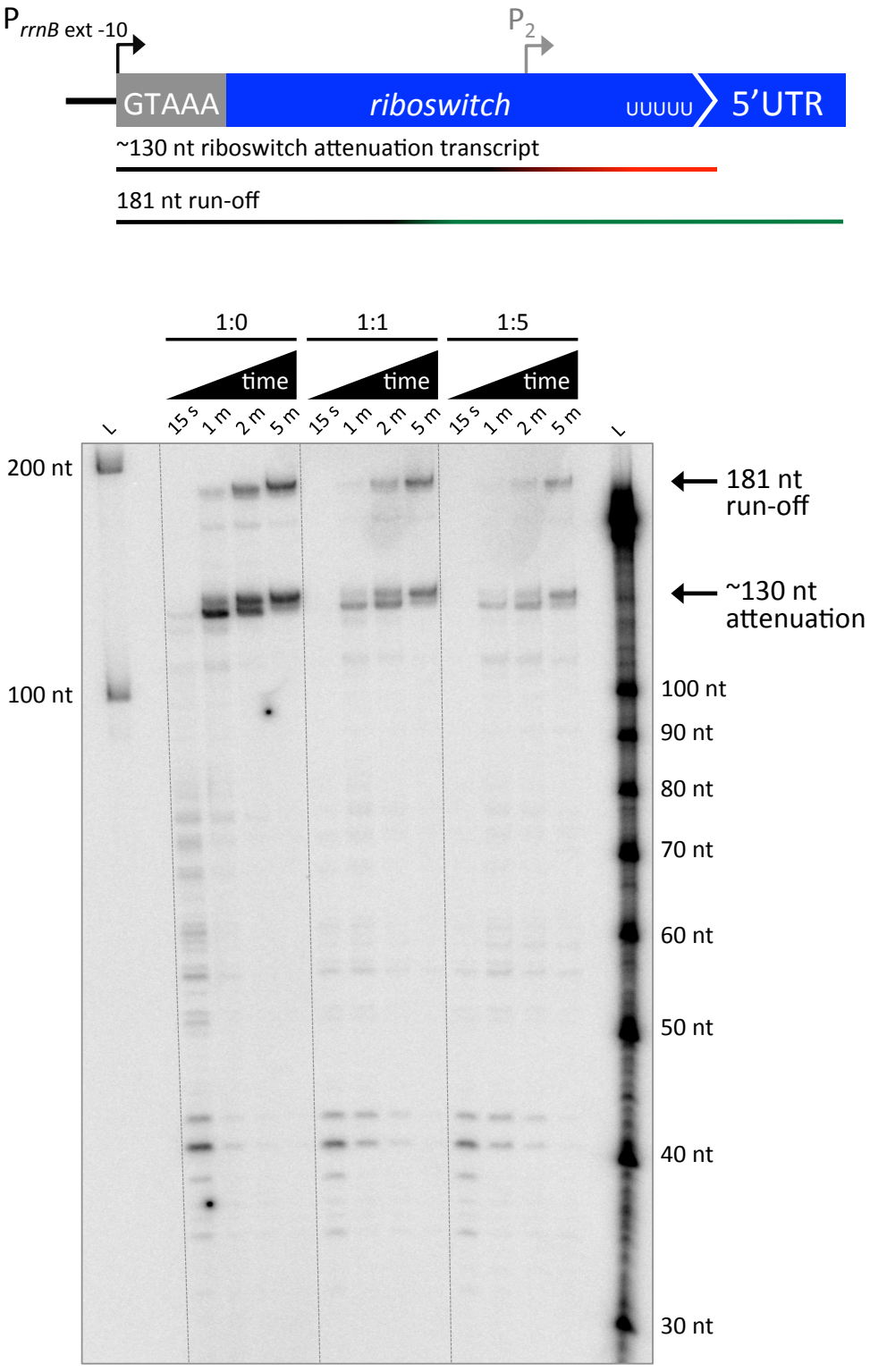


Figure 10.IX preliminary *in vitro* transcription with NusA

Single-round *in vitro* transcription time-course in the presence and absence of NusA. Reactions were carried out at 30°C to decrease processivity, with 17 nM *E. coli* RNAP (NEB), initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Reactions were carried out ± NusA (NEB) at molar ratios of 1:0, 1:1 and 1:5 relative to RNAP. Elongation was resumed with 25 µM NTPs, incubating for up to 5 min before stopping reactions with the addition of formamide loading dye. Template amplified with oligonucleotides 5.31 and 5.23 (see Appendix section 10.3).

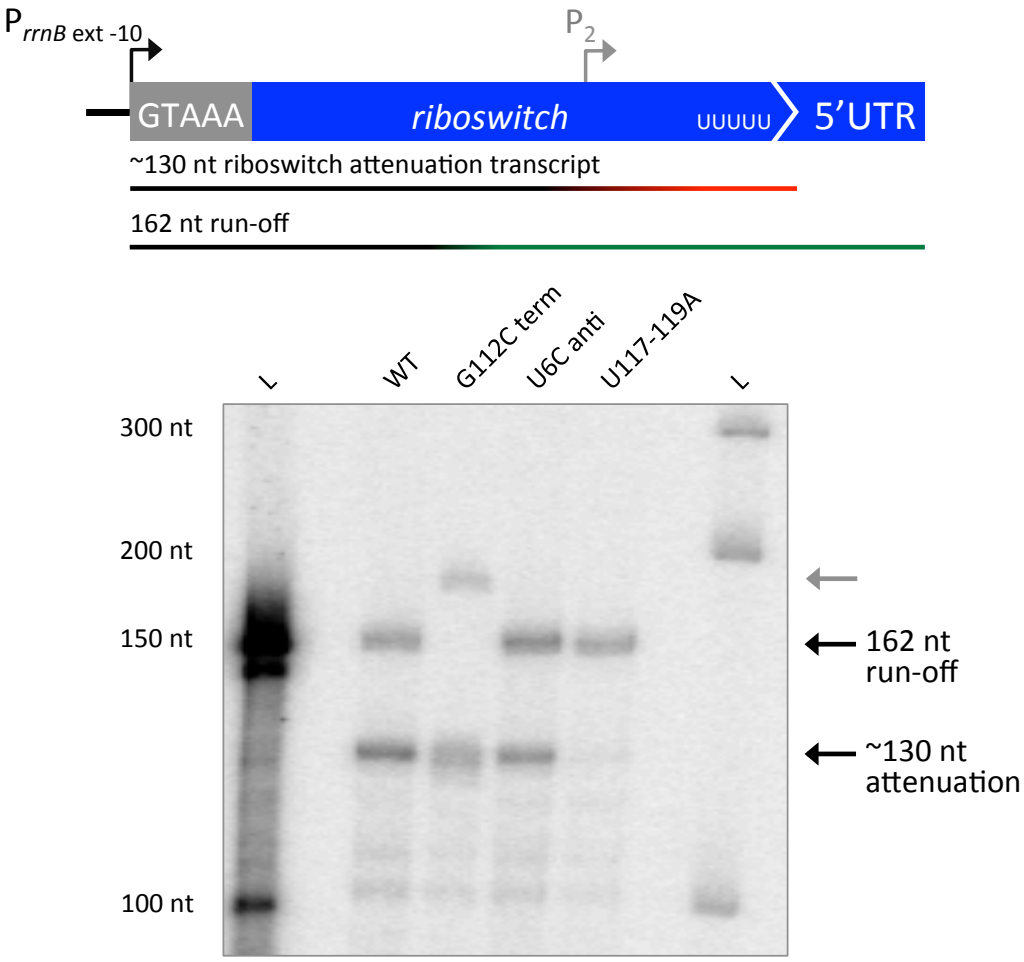


Figure 10.X 162 nt run-off template

Single-round *in vitro* transcription of riboswitch conformational mutants with 3'truncated template for 162 nt run-off Left to right: WT riboswitch template, G112C terminator stabilising substitution, U6C anti-terminator stabilising substitution and U117-119A intrinsic terminator disrupting substitution. Grey arrow denotes G112C transcription artefact. Reactions were carried out at 30°C to decrease processivity, with 17 nM *E. coli* RNAP (NEB), initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Elongation was resumed with 25 µM rNTPs, incubating for 20 min before stopping reactions with the addition of formamide loading dye. Template amplified with oligonucleotides 5.31 and 8.27 (see Appendix section 10.3).

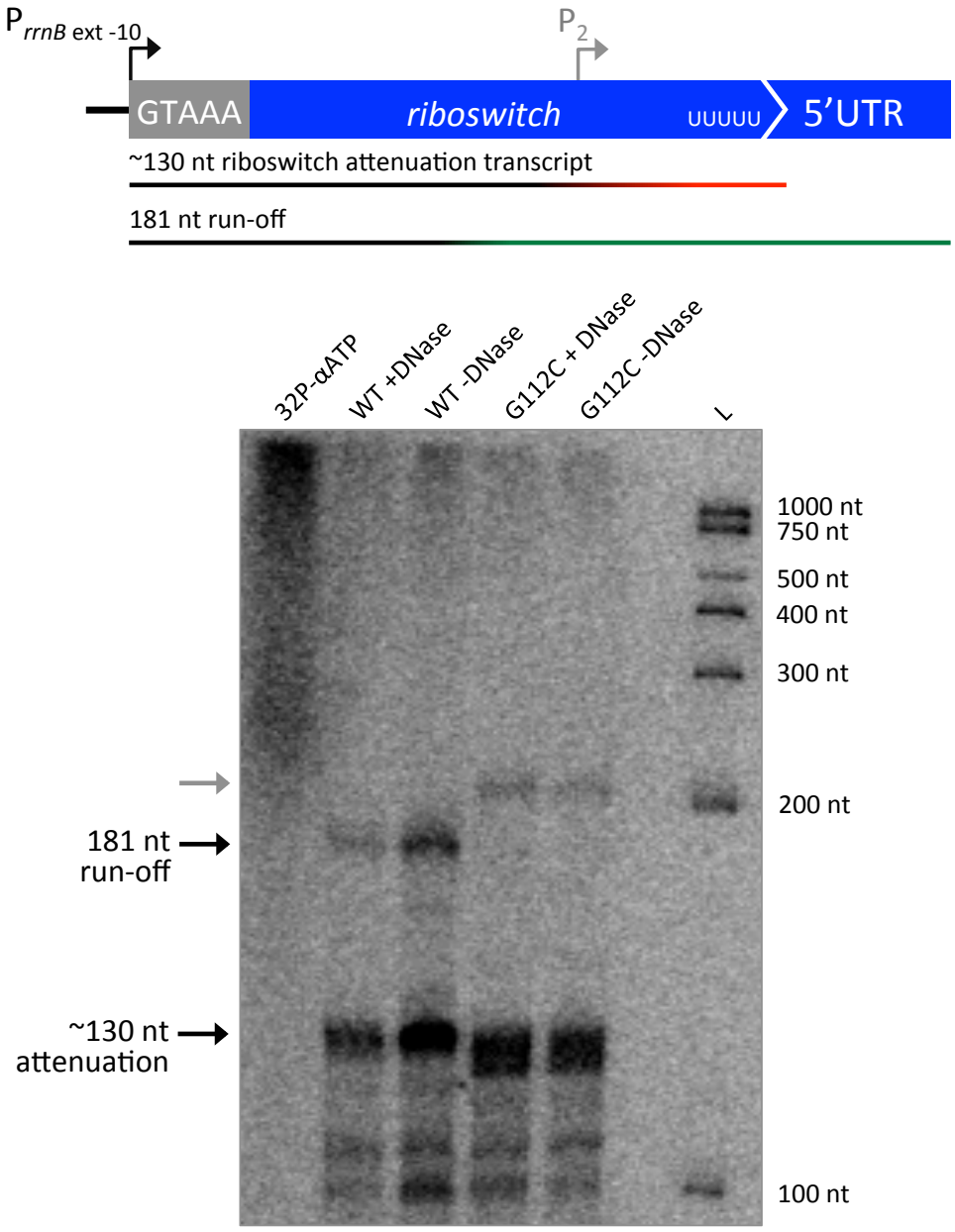


Figure 10.XI DNase treatment of *in vitro* transcription reactions
Single-round *in vitro* transcription of riboswitch subsequently treated with DNase I (Ambion). Left to right: WT riboswitch template with 32P-αUTP (Perkin Elmer) alone to encourage potential DNA labelling RNAP activity, WT template ± DNase treatment following 20 min elongation, G112C terminator stabilising substitution ±DNase treatment. Grey arrow denotes G112C transcription artefact. Reactions were carried out at 30°C to decrease processivity, with 17 nM *E. coli* RNAP (NEB), initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Elongation was resumed with 25 µM rNTPs, incubating for 20 min before stopping reactions by heating to 95°C for 5 min. Once cooled reactions were incubated with 1 µL DNase at 37°C for 30 min before addition of formamide loading dye. Template amplified with oligonucleotides 5.31 and 5.23 (see Appendix section 10.3).

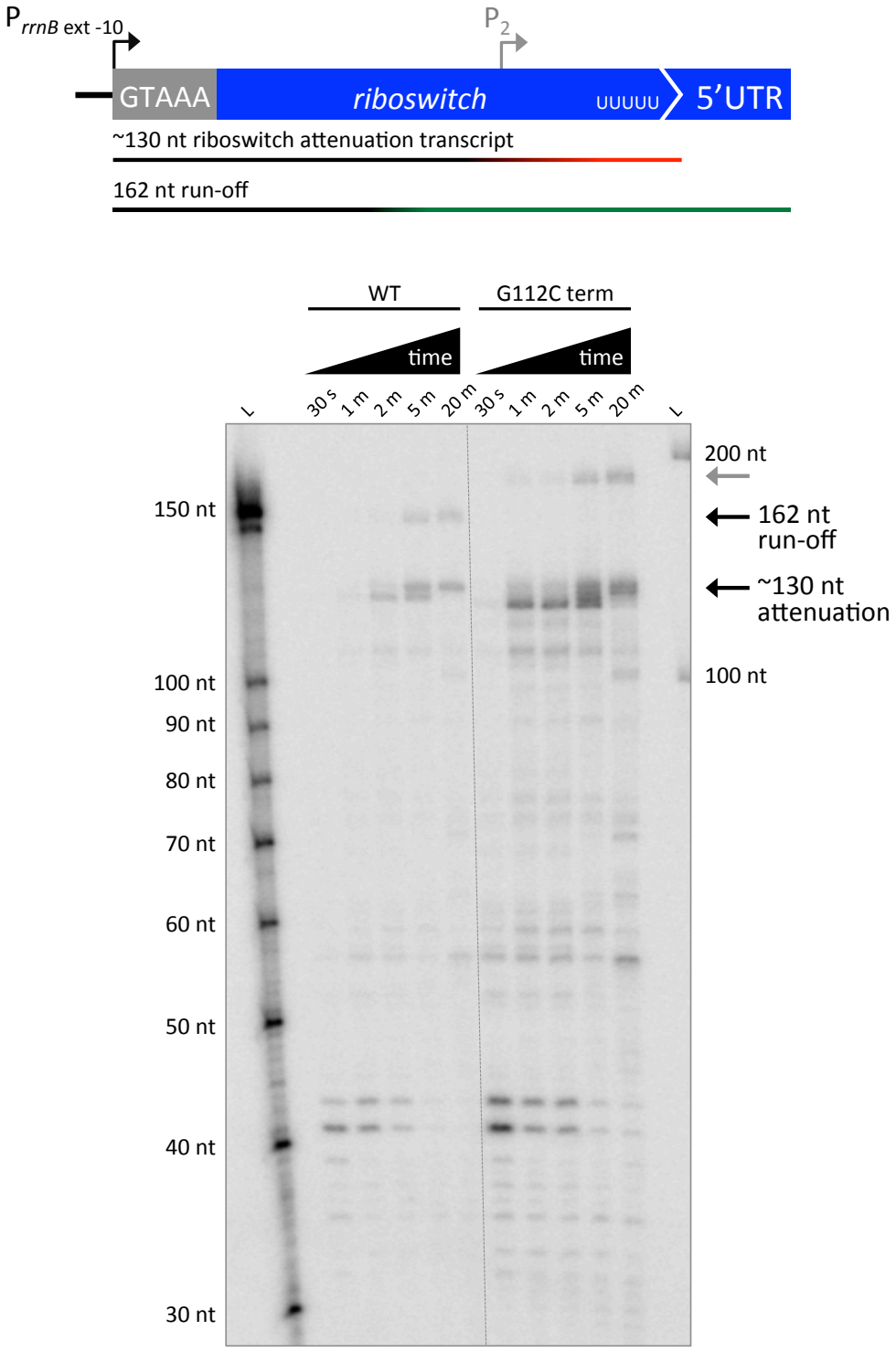


Figure 10.XII G112C *in vitro* transcription time-course
Single-round *in vitro* transcription time-course in the presence of NusA with 3'truncated template for 162 nt run-off for WT riboswitch template (left) and G112C terminator stabilising substitution (right). Grey arrow denotes G112C transcription artefact. Reactions were carried out at 30°C to decrease processivity, with 17 nM *E. coli* RNAP (NEB), initiated with GpU dinucleotide (IBA), labelled with 32P-αUTP (Perkin Elmer), limited to single-round transcription with 200 ng/µL heparin. Reactions were carried out in the presence of 85 nM NusA (NEB). Elongation was resumed with 25 µM NTPs, incubating for up to 20 min before stopping reactions with the addition of formamide loading dye. Template amplified with oligonucleotides 5.31 and 8.27 (see Appendix section 10.3).